Educational Achievement in Children:

Twinning, Teachers and Genes

Eveline de Zeeuw

READING COMMITTEE

prof.dr. C.V. Dolan	VU University Amsterdam
prof.dr. J. Jolles	VU University Amsterdam
prof.dr. P.F. de Jong	University of Amsterdam
dr. M.J.H. Rietveld	Windesheim Flevoland
prof.dr. P.C. Sham	University of Hong Kong
prof.dr. H.D. Webbink	Erasmus University Rotterdam

PARANYMPHS

Linda Bosschers
Vere van Koppen

ACKNOWLEDGEMENTS

We are grateful to the twin families and the teachers for their participation. This project is part of the research program 'Innovative learning materials and methods' funded by the Netherlands Initiative Brain and Cognition, a part of the Organization for Scientific Research (NWO) under grant 056-31-001. This research was supported by 'Spinozapremie' (NWO/SPI 56-464-14192); 'Twin-family database for behavior genetics and genomics studies' (NWO 480-04-004); Genetics of Mental Illness: European Research Council (ERC-230374); Genetic influences on stability and change in psychopathology from childhood to young adulthood (NWO/ZonMW 91210020); Genomics of Developmental Trajectories in Twins (ARRA RC2 2MH08995).

ISBN: 978-94-6259-562-0

Printed by: Ipskamp Drukkers, Amsterdam

VRIJE UNIVERSITEIT

Educational Achievement in Children:
Twinning, Teachers and Genes

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. F.A. van der Duyn Schouten,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Psychologie en Pedagogiek
op dinsdag 24 maart 2015 om 15.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Lia Eveline Jessica de Zeeuw

geboren te Rotterdam

promotoren: prof.dr. D.I. Boomsma
prof.dr. J.C.N. de Geus

copromotor: dr. C.E.M. van Beijsterveldt

TABLE OF CONTENTS

Chapter 1	General Introduction	7
Part I	*Risk Factors*	
Chapter 2	Twin specific risk factors in primary school achievements	13
Chapter 3	Do children perform and behave better at school when taught by same-gender teachers?	35
Part II	*Genetic Factors*	
Chapter 4	Arithmetic, reading and writing performance has a strong genetic component: A study in primary school children	53
Chapter 5	A review and meta-analysis of twin studies on educational achievement in primary school highlights the importance of genetic variation	83
Chapter 6	Childhood ODD and ADHD behavior: The effect of classroom sharing, gender, teacher gender and their interactions	129
Part III	*Molecular Genetics and Causality*	
Chapter 7	Polygenic scores associated with educational attainment in adults predict educational achievement and ADHD symptoms in children	165
Chapter 8	Longitudinal modeling suggests that ADHD symptoms lead to lower educational achievement in children	189
Chapter 9	Summary and General Discussion	207
Chapter 10	Nederlandse Samenvatting (Dutch Summary)	229
Appendices	Sample and Data Collection	239
	List of Publications	263
	Dankwoord	267

1.

GENERAL INTRODUCTION

Children differ in their ability to learn the subject material that is taught at school. Some master basic skills and pick up knowledge easily while others struggle to keep up with their peers. Educational achievement in children has several facets: it can be defined as the performance at school as assessed by the teacher, as the extent to which children achieve the educational goals corresponding to their grade level, or as the outcome on a standardized test at a particular age, such as the Dutch educational achievement test administered in the last grade of primary school (Cito, 2002). Even children of similar age, attending the same school and taught by the same teacher differ greatly in their performance at school. Low educational achievement is an important predictor of continued low achievement, school dropout, delinquency (Moilanen, Shaw & Maxwell, 2010) and of numerous other outcomes later in life, including lower earning (Julian & Kominski, 2011) and lower well-being (Mackenbach et al., 1997).

Numerous studies have also found a negative effect of attention deficit hyperactivity disorder (ADHD) and oppositional defiant disorder (ODD) on educational achievement (Greene et al., 2002; Polderman et al., 2010). Children with ADHD have difficulties with both inattention and hyperactivity or impulsiveness that interfere with daily functioning. ODD is characterized by hostile and defiant behavior towards figures with authority, going beyond normal childhood behavior (American Psychiatric Association, 2000). At school, children with ADHD have, for example, difficulty remaining in their seats and children with ODD often argue with their teachers. Children with ADHD and ODD receive lower grades and are more often referred to specialized education (Biederman et al., 1996; Greene et al., 2002).

The overarching aim of this thesis is to increase knowledge on the causes of individual differences in educational achievement and problem behavior of Dutch children attending primary school. This aim is to be achieved through a series of studies that are carried out based on data of young twins and their siblings. School performance was assessed by their teachers with the Teacher Report Form (TRF) (Achenbach, 1991), educational achievement was measured with objective standardized tests, pupil monitoring tests (Cito, 2014) for all grades and a national educational achievement test administered in the final grade (Cito, 2002) Problem behavior was rated by teachers with the short Conners' Teacher Rating Scales - Revised (CTRS-R) and by mothers with the short Conners' Parent Rating Scales - Revised (CPRS-R) (Conners et al., 1998; Conners, 2001). The objectives were threefold. First, to investigate the influence of twin specific risk factors on educational achievement by examining differences in educational achievement between twins and their non-twin siblings. Second, to determine the causes of individual differences in educational achievement and behavioral problems in primary school aged children by applying the classical twin design, which compares the resemblance in mono- and dizygotic twin pairs (Plomin et al., 2008). Third, to test the association between behavioral problems and educational achievement by using molecular genetic approaches and causality models.

Chapter 2 looks at the influence of several twin specific risk factors on school performance, the proficiency of a child as rated by their teacher in arithmetic, language, reading and physical education at the beginning of primary school and the educational achievement test in the final grade. The studied risk factors are not unique to twins, except for zygosity, but show a higher prevalence in twins than in singletons. The school performance of twins is compared to that of their non-twin siblings in a within-family design, thereby taking into account confounding of multiple demographic characteristics.

Chapter 3 describes the influence of an environmental factor, a same-gender teacher versus a different gender teacher, on school performance, educational achievement, and ADHD behavior of 12-year-old children in two genetically sensitive designs using monozygotic twin pairs who are discordant for the gender of their teacher(s) and dizygotic twin pairs of opposite-sex who are concordant for the gender of their teacher(s). Differences within these twin pairs are likely to be ascribed to the influence of the gender of the teacher, since (part of) their genotype, family background, social economic status, and multiple other characteristics of the twins are similar and thus controlled for. This chapter aims to contribute to the ongoing discussion in society as to whether children, especially boys, might be disadvantaged by the feminization of primary education.

Chapter 4 reports on the influence of genetic and environmental effects on educational achievement across the primary school years. Data on educational achievement were collected from teachers by asking them to send in a student report with the results of pupil monitoring tests assessing arithmetic, reading, reading comprehension and spelling from grades 1 to 6 (Cito, 2014). In the Netherlands, in addition to the pupil monitoring tests, a standardized educational achievement test is administered in the final grade. Possible gender differences in the etiology of educational achievement are also explored.

Chapter 5 provides a comprehensive review of the existing twin studies on educational achievement in primary school children (6-13 years) assessed in unselected genetically informative samples. To enhance statistical power and to estimate heritability across multiple data sets from the world literature, meta-analyses of twin correlations were performed for scores in several educational domains, i.e. reading, reading comprehension, mathematics and spelling. It was tested whether there were differences in the heritability of educational achievement between countries.

Chapter 6 looks at the extent to which individual differences in ODD and ADHD behavior are influenced by genetic effects and determines the moderation of the heritability by classroom sharing, gender of the student and gender of the teacher. ODD and ADHD behavior were assessed by the four scales of the CTRS-R. First, a series of models were investigated to test for measurement invariance (MI), across gender of the student and teacher for the four CTRS-R scales. MI means that children with the same (unobserved) ADHD or ODD vulnerability have the same probability of a response to a diagnostic item regardless of other characteristics of the child, such as its gender.

Chapter 7 examines whether the genetic variants, in this case genotyped single nucleotide polymorphisms (SNPs), that are associated with educational attainment in adults, are also associated with school performance and educational achievement in children. The first study of SNPs associated with educational attainment in adults was reported in 2013 in a publication in Science (Rietveld et al., 2013). The effect sizes from this study were used to calculate polygenic scores and to compute the explained variance in school performance and educational achievement in sample of 12-year-olds. Next, it was tested whether the polygenic scores for educational achievement also have an effect on ADHD, as rated by mothers and teachers, thereby testing if the relation between ADHD and educational achievement can (partly) be explained through genetic pathways.

Chapter 8 tests whether the negative association between ODD and ADHD behavior and educational achievement can be explained by a causal effect or genetic pleiotropy only. These tests were done by analyzing data from monozygotic and dizygotic twins whose ODD and ADHD behavior was assessed by their mothers, at the ages 7 and 12 years. Data for educational achievement came from the standardized educational achievement test administered in the last grade of primary school (age 12).

This thesis is concluded with a summary of the main results and a general discussion (**Chapter 9**) and a Dutch summary (**Chapter 10**). The data collection procedures, study sample and measurement instruments used in this thesis are described in a series of **Appendices**.

Part I

RISK FACTORS

2.

TWIN SPECIFIC RISK FACTORS IN PRIMARY SCHOOL ACHIEVEMENTS

Based on Eveline L. de Zeeuw, Catherina E. M. van Beijsterveldt, Eco J. C. de Geus and Dorret I. Boomsma (2011). Twin Specific Risk Factors in Primary School Achievements. *Twin Research and Human Genetics, 15 (1),* p. 107-115.

The main aim of this study was to examine twin specific risk factors that influence educational achievement in primary school. We included prenatal factors that are not unique to twins, except for zygosity, but show a higher prevalence in twins than in singletons. In addition, educational achievement was compared between twins and their non-twin siblings in a within-family design. Data were obtained from parents and teachers of approximately 10,000 twins and their non-twin siblings registered with the Netherlands Twin Register. Teachers rated the proficiency of the children on arithmetic, language, reading, and physical education, and reported a national educational achievement test score. Structural equation modeling showed that gestational age, birth weight, and sex were significant predictors of educational achievement, even after correction for socioeconomic status. Mode of delivery and zygosity did not have an effect, while parental age only influenced arithmetic. Mode of conception, incubator time, and birth complications negatively affected achievement in physical education. The comparison of educational achievement of twins and singletons showed significantly lower ratings on arithmetic, reading, and language in twins, compared to their older siblings, but not compared to their younger siblings. Low birth weight and small for gestational age were the most important risk factors for lower educational achievement of twins in primary school. It seems that the differences observed between twins and their non-twin siblings in educational achievement can largely be explained by birth order within the family.

INTRODUCTION

In the last decade, the twin birth rate was more than 32 per 1000 live births and twins are currently estimated to make up almost 2 per cent of the world population (Martin et al., 2010). Sharing the womb with another fetus can influence prenatal as well as perinatal conditions and outcomes. Fetuses have to compete for nutrition and, near the end of the pregnancy, for the best position in the uterus (Powers & Kiely, 1994). Crowding is a major risk factor for early birth and, as a consequence, twins are born, on average, 3 weeks earlier and with lower birth weights than singletons (Gielen et al., 2010). Second-born twins seem to suffer more from the sharing of the womb (Prins, 1994). For example, they have even lower mean birth weights than first-born twins (van Baal & Boomsma, 1998). After birth, most parents of twins perceive the first years when caring for two newborns as stressful and exhausting (Hay & O'Brien, 1984). As a consequence, twins may become each other's competitor when parents have to divide their attention between them. For example, it appears that mothers speak less often directly to one of the twins as an individual (Rutter & Redshaw, 1991). The difference in intrauterine environment and the limited resources in the family environment could influence the development of twins. Twin status is associated with several other potential risk factors for cognitive development, including assisted conception, prematurity, low birth weight, cesarean section, time spent in an incubator, and birth complications. These factors are not unique to twins, but are more prevalent in twins than in singletons, while risks associated with zygosity are specific to twins.

Almost 16 per cent of twins are born after assisted reproductive therapies (ART), compared to approximately 1 per cent of singletons (Wright et al., 2008). This difference is due to both the fertility drugs that increase the chance of several eggs being released at the same time and the practice of implanting more than one embryo, which was common in the early years of in vitro fertilization (IVF). A review of studies about the outcomes of children born after ART concluded that their cognitive development is comparable to that of naturally conceived children (Wilson et al., 2011). Another review also concluded that there is no evidence for differences in educational achievement between children born after assisted conception and children born after natural conception (Wagenaar et al., 2008).

More than half of all twins are born premature (gestational age of less than 37 weeks) and have low birth weight, according to the definition of the World Health Organization (less than 2500 grams), compared to less than 10 per cent of singletons (Martin et al., 2010). As a consequence, twins are placed in an incubator more often and for a longer period of time. However, the average difference in birth weight between twins and singletons is over 1000 grams (De Geus et al., 2001), and whether growth retardation is the same in twins as in singletons is still unknown (Phillips, Davies & Robinson, 2001). It seems that, for academic performance, the relative birth weight of twins is more important than their absolute birth weight (Christensen et al., 2006). However, after correcting for several potential confounders, a relationship between low birth weight and lower IQ (Aylward, 2005; Matte et al., 2001; Shenkin, Starr & Deary, 2004) and poorer educational achievement (Lundgren & Tuvemo, 2008) has been found in singletons as well as in twins.

Complications occur more frequently during the birth of twins and a cesarean section is more common than for singletons. While the pressure on the brain of children born through vaginal delivery might have a negative effect on the child's brain, it has also been suggested that the exposure of the developing brain to anesthesia during cesarean delivery has a negative influence on cognitive development (Khadem & Khadivzadeh, 2010). Yet, learning disabilities occur just as often in children born after cesarean section as in children born after vaginal delivery (Sprung et al., 2009), and there appears to be no association between intelligence and mode of delivery (Khadem & Khadivzadeh, 2010).

Finally, some adverse effects occur more often in monozygotic twins than in dizygotic twins. For example, monozygotic twins were more likely to have low birth weight or to be born preterm (Gielen et al., 2010; Hoskins, 1995). However, according to a meta-analysis, the difference in intelligence between twins and singletons was not influenced by zygosity status (Voracek & Haubner, 2008). Whether this also applies to educational achievement has not yet been determined. The question of whether twins differ from singletons in their cognitive abilities due to the risk factors associated with their twin status has been the focus of research for a long time (Hay & O'Brien, 1984; Record, McKeown & Edwards, 1970; Vandenberg, 1984). A recent meta-analysis of studies on differences in intelligence between twins and singletons concluded that, on average, twins seem to have lower IQs than singletons (Voracek & Haubner, 2008). The estimates of the difference in intelligence range from 5.1 to only 0.5 IQ points in studies from different countries, populations, and birth cohorts. Several studies based on birth cohorts from many years ago found a lower intelligence in twins (Deary et al., 2005; Record, McKeown & Edwards, 1970; Ronalds, de Stavola & Leon, 2005). However, prenatal and perinatal care has improved in the past decades, which may have reduced this difference in cognitive ability between twins and singletons. A study from the Netherlands found no evidence of differences in cognitive performance between adult twins and their non-twin siblings (Posthuma et al., 2000). A longitudinal study measured IQ scores in a large sample of singletons and approximately 6000 twins who went to primary school between 1994 and 2003. The study found that there was only a small difference (less than 1 IQ point) at ages 6 and 8 years, which disappeared at age 12 years (Webbink et al., 2008).

Intelligence is the single best predictor of educational achievement and correlates approximately 0.5 with school grades (Bartels et al., 2002). Therefore, when a difference in IQ is found between twins and singletons, the educational achievement of twins will probably be affected as well. Only a few studies have looked at the differences in educational achievement between twins and singletons. A 1983–1985 birth cohort study from Taiwan found that twins had lower scores and were less likely to attend college, even when the data were adjusted for birth weight, gestational age, birth order within the family, sex, and socioeconomic status (Tsou et al., 2008). The scores of Dutch female twins on an educational achievement test were also lower than those of singleton controls from the same grade and those of an older brother or sister. However, the twins performed just as well as the total Dutch female population and the difference found between twins and singletons was attributed to a bias in the selection of the control group (Cohen et al., 2002).

As with any phenotype, controlling for differences between twin and singleton families has been a problem in these studies. Selection bias, differences in social background, and family composition may explain differences between twins and singletons. Furthermore, most studies on the difference in educational achievement between twins and singletons have not corrected for the possible confounding influence of birth order within a family, which has been suggested to have an effect on intelligence. One study reported IQ to be approximately two IQ points lower in children with one older sibling (Bjerkedal et al., 2007). Another study found little effect of the number of older siblings on the difference in IQ between singletons and twins (Ronalds, De Stavola & Leon, 2005).

The present study used teacher ratings of different school subjects for twins in primary school from twin families on the Netherlands Twin Register. The data from teacher surveys on non-twin siblings of these twins provided a perfect match on social and family background. The first objective of our study was to determine the influence of several risk factors associated with twin birth on the educational achievement of twins. The second objective was to investigate whether the difference in intelligence found between twins and singletons also exists for educational achievement, taking into account the possibility that the birth or der of twins within a family may explain part of the difference in educational achievement between twins and singletons.

METHODS

PARTICIPANTS

The Netherlands Twin Register (NTR), established in 1987 by the Department of Biological Psychology at the VU University in Amsterdam, registers approximately 40 per cent of all multiple births in the Netherlands. The parents of these twins receive a survey about the development of their children every two years until the twins are 12 years old. At ages 7, 9, and 12 years, when the twins are attending primary school, parents are asked consent to approach the teacher(s) of their children with a survey. Since 2005, the siblings of twins in primary school are also included in the database (Bartels et al., 2007; Boomsma et al., 2002; Boomsma et al., 2006).

Information about the birth of the twins was obtained with the first survey sent to the parents shortly after registration of the newborns. This survey asks mothers to report on several birth characteristics, including maternal and paternal age at birth, mode of conception, gestational age, birth weight, time in an incubator, mode of delivery, birth complications, and sex. The educational achievement data were obtained with a survey sent to the primary school teachers. Teachers and parents were asked to report the scores of a national educational achievement test administered in the last grade of primary school (Cito, 2002).

The present study analyzed data from 7-year-old twins (*M* = 7.5, *SD* = .5) from birth cohorts 1992–2003 to determine the influence of twin and family risk factors on educational achievement in primary school (*N* = 9917). Questionnaires of children attending specialized education (*N* = 127) and questionnaires missing educational achievement data (*N* = 374) were excluded from this sample. The sample included data of children from 4272 complete twin pairs (*N* = 8544) and 872 twins from incomplete pairs. Incomplete data were due to one of the teachers not returning the questionnaire when the twins were in different classes or schools.

Because not all twins in the sample had reached the last grade of primary school yet, scores on the national test of educational achievement were not yet available for some of the twins. The data available in this sample included both teacher ratings and national educational achievement test score (*N* = 3262), only teacher ratings (*N* = 5944), or only a national educational achievement test score (*N* = 210). In this sample, 3012 twins belonged to an opposite-sex twin pair. For the twins belonging to a same-sex twin pair, determination of zygosity status was based on DNA polymorphisms (*N* = 603), on the first survey sent to the mother (*N* = 215), or on a parental questionnaire with 10 items about resemblance in appearance and frequency of mistaking the children for each other (*N* = 5530). With this last method, zygosity can be established with an accuracy of almost 93 per cent (Rietveld et al., 2000). Zygosity data was missing for 56 twins from 34 families. Information on the country of birth showed that 95.0% of mothers and 93.3% of fathers were born in the Netherlands, 1.4% and 1.9% in another Western country, and 1.5% and 2.3% in a non-Western country. For 1.9% of the mothers and 2.5% of the fathers, the country of birth was unknown.

For the analysis of the difference between twins (*M* = 9.1 years, *SD* = 1.8) and their non-twin siblings (*M* = 9.8 years, *SD* = 2.1), data from teacher surveys of 7, 9, and 12-year olds were analyzed. All twins for whom a teacher's survey of an additional non-twin sibling was available were included in the sample (*N* = 1375). This sample included 613 complete twin pairs (*N* = 1226), each pair matched with the non-twin sibling, as well as 149 twins from incomplete pairs and their non-twin siblings. Data on birth order within the family was available for 577 of the included families.

MEASURES

Educational achievement was assessed by the evaluation of several school subjects with two versions of a teacher's survey. In the first version of the survey (birth cohorts 1992–1998), teachers could choose up to six subjects and rated the proficiency of the students in these subjects on a five-point scale from 1 (*insufficient*) to 5 (*[very] good*). In the second version (birth cohorts 1997–2003), teachers rated the proficiency of the students in four predefined school subjects (arithmetic, language, reading, and physical education) on a similar five-point scale. Due to the free choice of school subjects in the first version and differences in missing data on the twin risk factors, the number of teacher ratings varies between the different comparison tests. The teacher survey also included questions about the type of education (regular or special) a child was attending, and whether he or she had ever had to repeat a grade.

The national test of educational achievement consists of multiple choice items in four different subjects (arithmetic, language, world studies [optional], and study skills), and is administered in the last grade of primary school. In this paper, the total score on the national educational achievement test, a standardized measure that ranges from 500 to 550, is used. The questions concerning world studies are not included in the total score since administration of these questions is optional.

Socioeconomic status (SES) was based on a full description of the occupations of both parents, and was coded according to the system of Statistics Netherlands (CBS Statistics, 2001), or an EPG-classification combined with information on parental education (Erikson, Goldthorpe & Portocarero, 2010). The SES score was classified on a five-point scale from 1 (*lower job*) to 5 (*scientific profession*), and the highest SES score of the parents determined the family SES. The SES score of the family when the twins were aged 3 or 10 years was used when the SES at age 7 years was not available, because these scores are highly correlated over time.

Assisted conception included in vitro fertilization (IVF) or intra cytoplasmic sperm injection (ICSI), and natural conception excluded conception after the prescription of ovulation-inducing drugs. Preterm birth was defined as born before 37 weeks gestation, and low birth weight was defined as less than 2500 grams. Birth complications were considered present when parents indicated that a child had experienced health problems directly after the delivery. Incubator time was defined as the number of days a child had spent in an incubator after delivery. Birth order within a family of twins and their non-twin sibling was determined on the basis of the order of the date of births in the families.

STATISTICAL ANALYSES

Data were analyzed by independent sample *t* tests to compare educational achievement between groups of twins that differed in mode of conception, gestational age, birth weight, incubator time, mode of delivery, birth complications, sex, and zygosity. Paired sample *t* tests were used to analyze the difference in educational achievement between twins and their non-twin siblings. A chi-square test was used to compare the number of grade repeaters amongst twins and their non-twin siblings. Data were checked and analyzed using the Statistical Package for the Social Sciences (SPSS 17.0) (2011). The tests were done in the statistical program Stata 9.0 (StataCorp, 2005) to correct for the influence of the family cluster effect. For all analyses, two-tailed *p* values of $< .05$ indicated statistical significance.

A linear structural equation model was estimated in Lisrel 8 (Jöreskog & Sörbom, 2002) to simultaneously investigate the influence of twin and family risk factors on educational achievement. The model included all twin and family risk factors as independent latent variables, and the teacher rating for the four school subjects as dependent variables. Correlations between the dependent variables were estimated. The analyses were based on the full information likelihood maximization. To correct for the family cluster effect, the data were divided into two groups. The first-born and second-born twins of every twin pair were randomly assigned to the first or second group. The Lisrel model was fitted to the data of both groups to determine whether the results found in the first group could be replicated in the second group.

TABLE 1 Twin and family risk factors

	N	(out of N families)	%
Socioeconomic Status			
1 Low	130	(75)	1.4
2	1113	(601)	11.8
3	4112	(2231)	43.7
4	2299	(1239)	24.4
5 High	1234	(681)	13.1
missing	*528*	*(300)*	*5.6*
Mode of Conception			
Natural	6828	(3715)	76.5
IVF/ICSI	1339	(733)	15.0
Ovulation Inducing Drugs	766	(415)	8.1
missing	*483*	*(268)*	*5.1*
Maternal Age			
<25 years	243	(111)	2.6
25-30 years	2564	(1383)	27.2
30-35 years	4430	(2419)	47.0
>35 years	2108	(1156)	22.4
missing	*71*	*(41)*	*0.8*
Paternal Age			
<25 years	76	(41)	0.8
25-35 years	5349	(2916)	56.8
35-45 years	3536	(1932)	37.6
>45 years	220	(121)	2.3
missing	*235*	*(132)*	*2.5*
Zygosity			
Monozygotic	3246	(1751)	34.5
Dizygotic	6114	(3346)	64.9
missing	*56*	*(34)*	*0.6*
Gestational Age			
>37 weeks	5570	(3030)	59.2
32-37 weeks	2988	(1624)	31.7
<32 weeks	660	(365)	7.0
missing	*198*	*(112)*	*2.1*

Birth Weight		
<1500 grams	414	4.4
1500-2500 grams	3840	40.8
>2500 grams	4881	51.8
missing	*281*	3.0
Incubator Time		
0 days	4638	49.3
1-7 days	2570	27.3
8-14 days	727	7.7
>14 days	942	10.0
missing	*539*	5.7
Mode of Delivery		
Vaginal	6442	68.4
Cesarean Section	2786	29.6
missing	*188*	2.0
Birth Complications		
No	7043	74.8
Yes	2086	22.2
missing	*287*	3.0
Sex		
Boy	4634	49.2
Girl	4782	50.8

RESULTS

Table 1 shows the summary statistics of the twin and family risk factors of the 7-year-old twins, and Table 2 displays the means of educational achievement for each risk factor. The analysis of several twin birth risk factors showed that twins born after assisted conception had significantly higher teacher ratings for reading and significantly lower performance in physical education, compared to twins born after natural conception. However, when matched on the possible confounders SES, maternal age at birth, and birth order within a family, twins born after assisted conception were no longer better at reading ($t = -0.55$, $p = .585$). Their achievement in physical education remained lower ($t = 2.14$, $p = .033$). Preterm twins, twins who had to be placed in an incubator, and twins with complications after delivery had poorer performance in physical education. Low birth weight twins received lower ratings for arithmetic and physical education and scored lower on the national educational achievement test. Mode of delivery had no effect on any of the school subjects. There were significant sex differences: boys performed better at arithmetic and obtained higher scores for the educational achievement test, while girls received higher ratings for language, reading, and physical education. No significant differences were found between the educational achievement of monozygotic and dizygotic twins.

TABLE 2 Means of educational achievement for the twin risk factors

	N	Mean	N	Mean	t	p
Mode of Conception	**Natural**		**Assisted**			
Arithmetic	6504	3.76	1274	3.81	-1.28	.202
Language	4615	3.67	950	3.65	.47	.639
Reading	5640	3.49	1109	3.58	-2.05	.041*
Physical Education	3401	3.85	745	3.76	2.24	.025*
Educational Achievement	2576	537.6	458	538.3	-1.48	.139
Gestational Age	**Full Term**		**Preterm**			
Arithmetic	5288	3.80	3483	3.76	1.35	.178
Language	3823	3.69	2459	3.67	.63	.528
Reading	4593	3.52	3012	3.52	.08	.940
Physical Education	2776	3.86	1891	3.77	3.04	.002**
Educational Achievement	2113	537.5	1346	538.0	-1.27	.205
Birth Weight	**Normal**		**Low**			
Arithmetic	4655	3.84	4037	3.72	-4.69	<.001**
Language	3359	3.70	2851	3.66	-1.40	.161
Reading	4034	3.54	3498	3.50	-1.21	.227
Physical Education	2460	3.87	2169	3.77	-3.66	<.001**
Educational Achievement	1871	538.1	1563	537.2	-2.88	.004**
Incubator	**No**		**Yes**			
Arithmetic	4430	3.81	4320	3.76	1.89	.059
Language	3251	3.69	3002	3.67	.80	.421
Reading	3855	3.53	3725	3.52	.52	.604
Physical Education	2385	3.87	2269	3.78	3.17	.002**
Educational Achievement	1686	538.0	1757	537.4	1.87	.061
Mode of Delivery	**Vaginal**		**Cesarean Section**			
Arithmetic	6146	3.79	2636	3.77	.76	.445
Language	4343	3.68	1949	3.68	-.11	.916
Reading	5310	3.51	2304	3.56	-1.54	.124
Physical Education	3139	3.83	1534	3.81	.76	.450
Educational Achievement	2501	537.5	950	538.1	-1.57	.117

Birth Complications	**No**		**Yes**			
Arithmetic	6702	3.78	1983	3.80	-1.02	.308
Language	4823	3.68	1391	3.71	-.90	.370
Reading	5826	3.52	1699	3.53	-.22	.826
Physical Education	3514	3.87	1103	3.70	5.26	<.001**
Educational Achievement	2645	537.6	781	537.7	-.22	.826
Sex	**Boy**		**Girl**			
Arithmetic	4422	3.89	4539	3.66	9.47	<.001**
Language	3154	3.59	3300	3.77	-6.55	<.001**
Reading	3817	3.42	3952	3.62	-6.27	<.001**
Physical Education	2453	3.80	2380	3.86	-2.45	.015*
Educational Achievement	1638	538.3	1834	537.1	3.89	<.001**
Zygosity	**Dizygotic**		**Monozygotic**			
Arithmetic	5842	3.79	3066	3.75	1.44	.149
Language	4233	3.68	2170	3.67	.31	.754
Reading	5003	3.52	2721	3.51	.28	.783
Physical Education	3144	3.82	1637	3.85	-1.13	.259
Educational Achievement	2211	537.6	1261	537.8	-.43	.667

* p<0.05; ** p<0.01

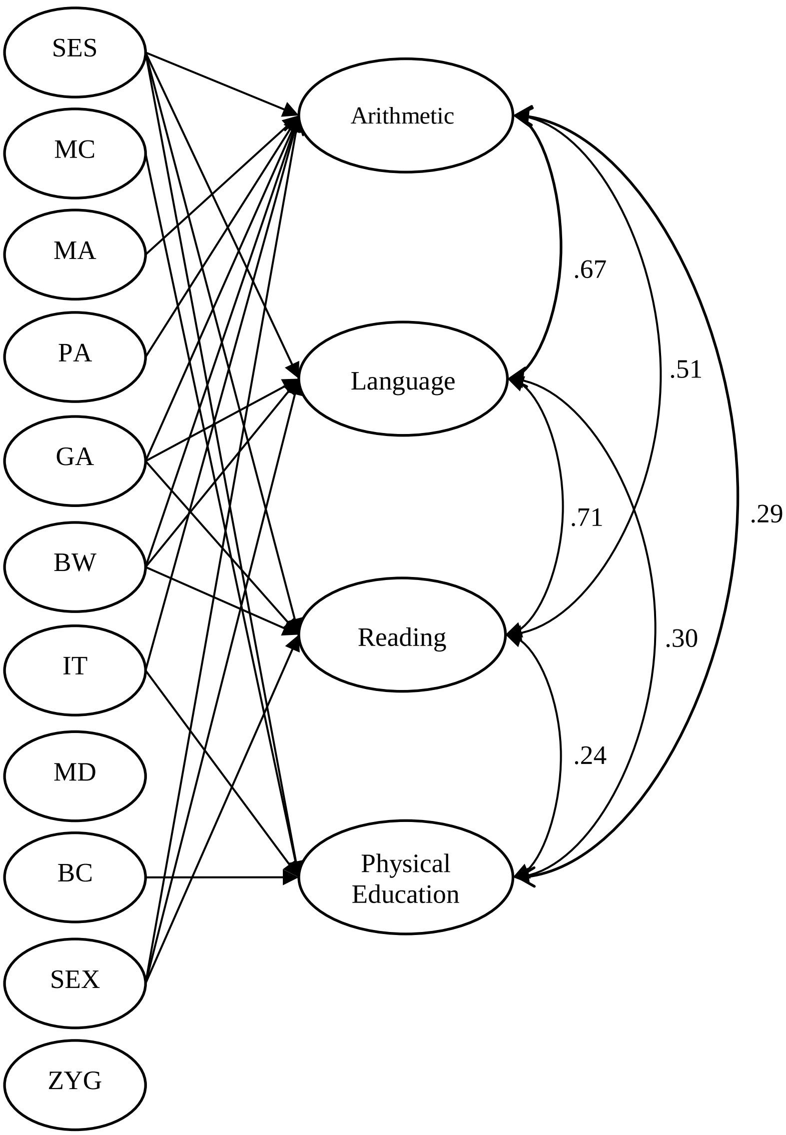
Figure 1 depicts the Lisrel model, with the eleven variables that represent the twin and family risk factors on the left and, on the right, the four school subjects as dependent factors. The twin and family risk factors together explained 3.8, 3.3, 2.7, and 2.5 (Group 1) per cent and 4.2, 2.9, 2.7, and 1.6 (Group 2) per cent of the variance, in arithmetic, language, reading, and physical education, respectively. Table 3 shows the parameter estimates and standard errors for the linear relationships between those independent and dependent variables. Socioeconomic status had a significant influence on the ratings of all four school subjects, and there were sex differences for arithmetic, language, and reading. Maternal age had a positive influence on the performance in arithmetic, while paternal age had a negative influence on this subject. Birth weight and gestational age influenced the ratings of all school subjects, except for physical education. Physical education was affected by mode of conception, incubator time, and birth complications. Mode of delivery and zygosity did not have an effect on educational achievement.

TABLE 3 Parameter estimates of the linear relationships between twin and family risk factors and educational achievement

	Arithmetic		Language		Reading		Physical Education	
	Parameter Estimate	SE	Parameter Estimate	SE	Parameter Estimate	SE	Parameter Estimate	SE
Socioeconomic status (SES)	.107**[1]	.011	.108**[1]	.012	.111**[1]	.014	.037**	.012
Mode of conception (MC)	.037	.048	.011	.050	.078	.059	-.106*	.052
Maternal age (MA)	.013*[1]	.013	.005	.006	-.001	.007	.003	.006
Paternal age (PA)	-.008[2]	.004	-.006	.005	-.003	.005	-.001	.005
Gestational age (GA)	-.014[2]	.011	-.022*	.011	-.035**	.013	.002	.011
Birth weight (BW)	.090*[1]	.044	.108*[1]	.047	.163**	.054	-.033	.048
Incubator time (IT)	-.004[2]	.002	-.001	.003	.001	.003	-.007**	.003
Mode of delivery (MD)	-.019	.034	-.004	.035	.047	.041	-.030	.036
Birth complications (BC)	.068	.039	.034	.040	-.026	.047	-.147**	.042
Sex (SEX)	-.202**[1]	.031	.200**[1]	.032	.215**[1]	.038	.058	.033
Zygosity (ZYG)	.025	.034	.031	.036	.036	.042	-.021	.037

note: [1] significance replicated in second group; [2] relationship significant in second group
* $p<0.05$; ** $p<0.01$

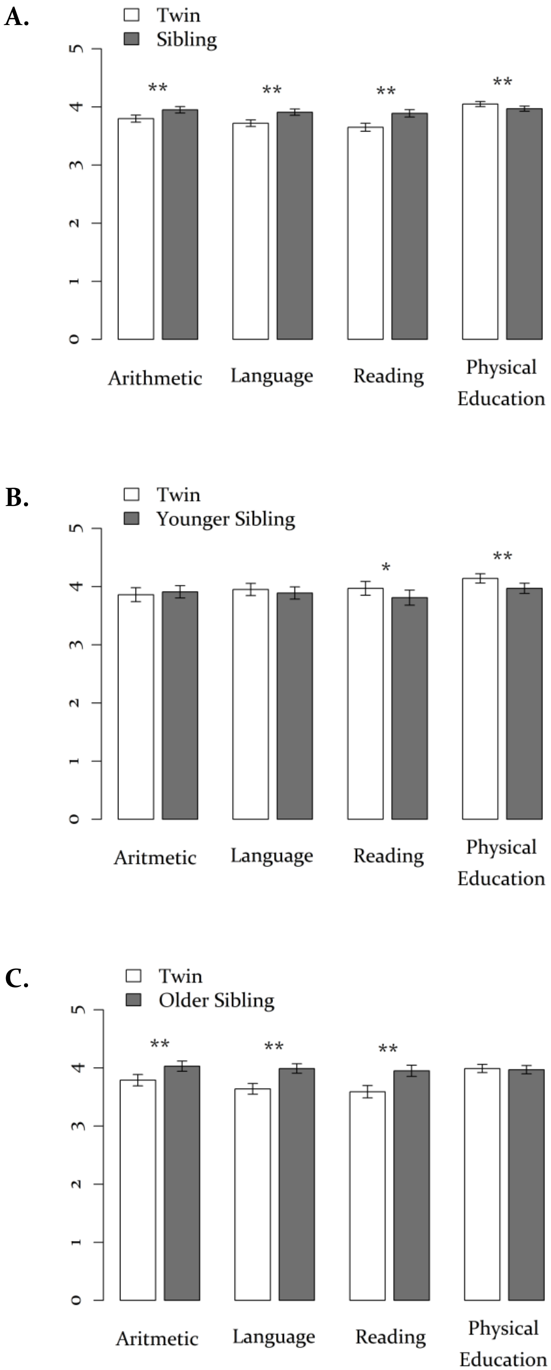
FIGURE 1 Structural equation model with the significant linear relationships between twin and family risk factors and educational achievement



SES = socioeconomic status; MC = mode of conception; MA = maternal age; PA = paternal age; GA = gestational age; BW = birth weight; IT = incubator time; MD = mode of delivery; BC = birth complications; SEX = sex; ZYG = zygosity

Twins were matched with a non-twin sibling to determine the difference in educational achievement between twins and singletons (N = 1375). The results show that singletons had significantly higher ratings in arithmetic (M_{twin} = 3.80, M_{sib} =3.95, t = −4.08, p < .001), language (M_{twin} = 3.72, M_{sib}= 3.91, t = −5.60, p < .001), and reading (M_{twin} = 3.65, M_{sib} = 3.89, t = −5.77, p < .001) (Figure 2a). However, 148 of 1367 (10.8%) twins were held back a year at some point in their school career, compared to 67 of 747 (10.1%) singletons ($\chi 2$ = 1.82, p = .177). In order to test whether the difference in educational achievement could be explained by the birth order of the twins within the family, separate analyses were performed on a group of twins who were first-born and a group of twins for whom the non-twin sibling was the first-born within the family. The results showed that twins with a younger sibling had the same, or even higher, ratings on arithmetic, language, and reading as their non-twin sibling (Figure 2b). In contrast, twins with an older sibling had significantly lower ratings than their non-twin sibling for these school subjects (Figure 2c). Physical education was an exception because all twins received higher ratings for this school subject than their non-twin sibling (M_{twin} = 4.05, M_{sib} = 3.97, t = 2.74, p = .007).

FIGURE 2 Differences in educational achievement between twins and their non-twin siblings



DISCUSSION

This study showed that gestational age and birth weight were the most important risk factors in twins. Twins with lower birth weight and small for gestational age performed more poorly in arithmetic, language, reading, and a national educational achievement test. Incubator time and paternal age had a negative effect on the ratings in arithmetic, while maternal age had a positive influence on this school subject. Achievement in physical education was negatively affected by mode of conception, incubator time, and birth complications, even after correction for gestational age and birth weight. The other risk factors, mode of delivery and zygosity, had no effect on educational achievement.

In agreement with IQ studies amongst preterm and low birth weight singletons and twins (Kirkegaard et al., 2006; Shenkin, Starr & Deary, 2004), birth weight had a negative effect on the educational achievement of twins in primary school. Assisted conception does not affect the educational achievement of twins once SES, maternal age at birth, and birth order are taken into account. This is in agreement with a study that found that children born after IVF scored even higher than matched controls on an achievement test (Mains et al., 2010). However, achievement in physical education was lower in children born after IVF/ICSI, compared to children born after natural conception. Our study is the first to investigate whether mode of delivery has an influence on educational achievement, and found that twins born after cesarean section have the same ratings on all school subjects as twins born after vaginal delivery. Finally, in agreement with studies that found that zygosity status does not influence intelligence (Voracek & Haubner, 2008), our study also shows that educational achievement did not differ between monozygotic and dizygotic twins.

When interpreting these results, one must keep in mind that the twin and family factors included in the Lisrel model explained only a small amount of the variance in educational achievement. The control variables socioeconomic status and sex had the greatest influence on the teacher ratings of the different school subjects. In addition, not all significant linear relationships between the independent and dependent variables in the Lisrel model were replicated in the second group. Of note, socioeconomic status and sex, here used as covariates, had a much greater influence on the teacher ratings of the different school subjects than the twin-specific risk factors of interest.

A difference between twins and singletons in educational achievement in primary school seems to exist in this sample. Singletons received higher ratings from their teachers for arithmetic and language, as well as for reading. However, an important observation is that birth order within the family can largely account for the lower educational achievement of twins found in this sample: twins who were first in birth order within the family had the same, or even higher, ratings as their non-twin sibling, while twins with a sibling who was first in birth order within the family had lower ratings than their non-twin siblings. Remarkably, all twins had a somewhat higher score on physical education, compared to their non-twin siblings.

Regarding the difference in educational achievement between twins and singletons, the results in the literature are mixed. One study also reported a difference in educational achievement of twins compared to singletons (Tsou et al., 2008). In contrast, a study from Denmark in birth cohorts from 1986 to 1988 showed that, although twins had, on average, a lower birth weight than singletons, their mean scores on a test of general academic achievement were as high as scores from singletons (Christensen et al., 2006). However, the study in Denmark used a random sample of singletons from the general population as a control group, while our study compared the twins with their non-twin siblings. The lower educational achievement of twins compared to singletons found in this study is also in accordance with a meta-analysis that concluded that an IQ difference exists between twins and singletons for multiple birth cohorts from various countries (Voracek & Haubner, 2008).The finding that birth order within the family could explain part of the differences in educational achievement between twins and their non-twin siblings is in agreement with another study in a group of twins. This study showed that intelligence was negatively associated with birth order within the family. Twins without older siblings had the highest IQs, while twins with two or more older siblings had the lowest IQs (Boomsma et al., 2008).

To conclude, our study is the first to give an overview of the influence of several risk factors associated with twin birth, including mode of conception, gestational age, birth weight, incubator time, mode of delivery, birth complications, and zygosity, on the educational achievement of a very large sample of 7-year-old twins in primary school. Low birth weight is the most important risk factor for the educational achievement of twins in primary school. The differences in educational achievement observed between 7, 9, and 12-year-old twins and their non-twin siblings can largely be explained by birth order within the family.

REFERENCES

IBM Corp (2011). IBM SPSS Statistics for Windows Version 20.0. Armonk, NY: IBM Corp.

Aylward, G.P. (2005). Neurodevelopmental outcomes of infants born prematurely. *Journal of Developmental and Behavioral Pediatrics, 26 (6)*, p. 427-440.

Bartels, M., Rietveld, M.J., van Baal, G.C., & Boomsma, D.I. (2002). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Research, 5 (6)*, p. 544-553.

Bartels, M., van Beijsterveldt, C.E., Derks, E.M., Stroet, T.M., Polderman, T.J., Hudziak, J.J., & Boomsma, D.I. (2007). Young Netherlands Twin Register (Y-NTR): a longitudinal multiple informant study of problem behavior. *Twin Resarch and Human Genetics, 10 (1)*, p. 3-11.

Bjerkedal, T., Kristensen, P., Skjeret, G.A., & Brevik, J.I. (2007). Intelligence test scores and birth order among young Norwegian men (conscripts) analyzed within and between families. *Intelligence, 35 (5)*, p. 503-514.

Boomsma, D.I., de Geus, E.J., Vink, J.M., Stubbe, J.H., Distel, M.A., Hottenga, J.J., Posthuma, D., van Beijsterveldt, T.C., Hudziak, J.J., Bartels, M., & Willemsen, G. (2006). Netherlands Twin Register: from twins to twin families. *Twin Resarch and Human Genetics, 9 (6)*, p. 849-857.

Boomsma, D.I., Vink, J.M., van Beijsterveldt, T.C., de Geus, E.J., Beem, A.L., Mulder, E.J., Derks, E.M., Riese, H., Willemsen, G.A., Bartels, M., van den, B.M., Kupper, N.H., Polderman, T.J., Posthuma, D., Rietveld, M.J., Stubbe, J.H., Knol, L.I., Stroet, T., & van Baal, G.C. (2002). Netherlands Twin Register: a focus on longitudinal research. *Twin Research, 5 (5)*, p. 401-406.

Boomsma, D.I., van Beijsterveld, T.C., Beem, A.L., Hoekstra, R.A., Polderman, T.J., & Bartels, M. (2008). Intelligence and birth order in boys and girls. *Intelligence, 36 (6)*, p. 630-634.

CBS Statistics (2001). Standard classification of occupations.

Christensen, K., Petersen, I., Skytthe, A., Herskind, A.M., McGue, M., & Bingley, P. (2006). Comparison of academic performance of twins and singletons in adolescence: follow-up study. *British Medical Journal, 333 (7578)*, p. 1095.

Cito (2002). Eindtoets Basisonderwijs. Arnhem, The Netherlands: Cito.

Cohen, C.C., Van Goozen, S.H., Orlebeke, J.F., Buitelaar, J.K., & Cohen-Kettenis, P.T. (2002). A comparison of educational achievement in a national sample of Dutch female twins and their matched singleton controls. *Twin Research, 5 (4)*, p. 273-276.

De Geus, E.J., Posthuma, D., IJzerman, R.G., & Boomsma, D.I. (2001). Comparing blood pressure of twins and their singleton siblings: being a twin does not affect adult blood pressure. *Twin Research, 4 (5)*, p. 385-391.

Deary, I.J., Pattie, A., Wilson, V., & Whalley, L.J. (2005). The cognitive cost of being a twin: two whole-population surveys. *Twin Research and Human Genetics, 8 (4)*, p. 376-383.

Erikson, R., Goldthorpe, J.H., & Portocarero, L. (2010). Intergenerational class mobility and the convergence thesis: England, France and Sweden. 1979. *British Journal of Sociology, 61 (S1)*, p. 185-219.

Gielen, M., van Beijsterveldt, C.E., Derom, C., Vlietinck, R., Nijhuis, J.G., Zeegers, M.P., & Boomsma, D.I. (2010). Secular trends in gestational age and birthweight in twins. *Human Reproduction, 25 (9)*, p. 2346-2353.

Hay, D.A. & O'Brien, P.J. (1984). The role of parental attitudes in the development of temperament in twins at home, school and in test situations. *Acta Geneticae Medicae et Gemellologiae, 33 (2)*, p. 191-204.

Hoskins, R.E. (1995). Zygosity as a risk factor for complications and outcomes of twin pregnancy. *Acta Geneticae Medicae Gemellologiae, 44 (1)*, p. 11-23.

Jöreskog, K. & Sörbom, D. (2002). PRELIS 2 user's reference guide. Chicago Scientific Software.

Khadem, N. & Khadivzadeh, T. (2010). The intelligence quotient of school aged children delivered by cesarean section and vaginal delivery. *Iranian Journal of Nursing and Midwifery Research, 15 (3)*, p. 135-140.

Kirkegaard, I., Obel, C., Hedegaard, M., & Henriksen, T.B. (2006). Gestational age and birth weight in relation to school performance of 10-year-old children: a follow-up study of children born after 32 completed weeks. *Pediatrics, 118 (4)*, p. 1600-1606.

Lundgren, E.M. & Tuvemo, T. (2008). Effects of being born small for gestational age on long-term intellectual performance. *Best Practice and Research: Clinincal Endocrinology and Metabolism, 22 (3)*, p. 477-488.

Mains, L., Zimmerman, M., Blaine, J., Stegmann, B., Sparks, A., Ansley, T., & Van, V.B. (2010). Achievement test performance in children conceived by IVF. *Human Reproduction, 25 (10)*, p. 2605-2611.

Martin, J.A., Hamilton, B.E., Sutton, P.D., Ventura, S.J., Mathews, T.J., Kirmeyer, S., & Osterman, M.J. (2010). Births: final data for 2007. *National Vital Statistics Reports, 58 (24)*, p. 1-85.

Matte, T.D., Bresnahan, M., Begg, M.D., & Susser, E. (2001). Influence of variation in birth weight within normal range and within sibships on IQ at age 7 years: cohort study. *British Medical Journal, 323 (7308)*, p. 310-314.

Phillips, D.I., Davies, M.J., & Robinson, J.S. (2001). Fetal growth and the fetal origins hypothesis in twins - problems and perspectives. *Twin Research, 4 (5)*, p. 327-331.

Posthuma, D., de Geus, E.J., Bleichrodt, N., & Boomsma, D.I. (2000). Twin-singleton differences in intelligence? *Twin Research, 3 (2)*, p. 83-87.

Powers, W.F. & Kiely, J.L. (1994). The risks confronting twins: a national perspective. *American Journal of Obstetrics and Gynecology, 170 (2)*, p. 456-461.

Prins, R.P. (1994). The second-born twin: can we improve outcomes? *American Journal of Obstetrics and Gynecology, 170 (6)*, p. 1649-1656.

Record, R.G., McKeown, T., & Edwards, J.H. (1970). An investigation of the difference in measured intelligence between twins and single births. *Annals of Human Genetics, 34 (1)*, p. 11-20.

Rietveld, M.J., van, D., V, Bongers, I.L., Stroet, T.M., Slagboom, P.E., & Boomsma, D.I. (2000). Zygosity diagnosis in young twins by parental report. *Twin Research, 3 (3)*, p. 134-141.

Ronalds, G.A., De Stavola, B.L., & Leon, D.A. (2005). The cognitive cost of being a twin: evidence from comparisons within families in the Aberdeen children of the 1950s cohort study. *British Medical Journal, 331 (7528)*, p. 1306.

Rutter, M. & Redshaw, J. (1991). Annotation: growing up as a twin: twin-singleton differences in psychological development. *Journal of Child Psychology and Psychiatry, 32 (6)*, p. 885-895.

Shenkin, S.D., Starr, J.M., & Deary, I.J. (2004). Birth weight and cognitive ability in childhood: a systematic review. *Psychological Bulletin, 130 (6)*, p. 989-1013.

Sprung, J., Flick, R.P., Wilder, R.T., Katusic, S.K., Pike, T.L., Dingli, M., Gleich, S.J., Schroeder, D.R., Barbaresi, W.J., Hanson, A.C., & Warner, D.O. (2009). Anesthesia for cesarean delivery and learning disabilities in a population-based birth cohort. *Anesthesiology, 111 (2)*, p. 302-310.

StataCorp (2005). Stata Statistical Software. College Station, TX: StataCorp.

Tsou, M.T., Tsou, M.W., Wu, M.P., & Liu, J.T. (2008). Academic achievement of twins and singletons in early adulthood: Taiwanese cohort study. *British Medical Journal, 337* p. a438.

van Baal, C.G. & Boomsma, D.I. (1998). Etiology of individual differences in birth weight of twins as a function of maternal smoking during pregnancy. *Twin Research, 1 (3)*, p. 123-130.

Vandenberg, S.G. (1984). Does a special twin situation contribute to similarity for abilities in MZ and DZ twins? *Acta Geneticae Medicae Gemellologiae, 33 (2)*, p. 219-222.

Voracek, M. & Haubner, T. (2008). Twin-singleton differences in intelligence: a meta-analysis. *Psychological Reports, 102 (3)*, p. 951-962.

Wagenaar, K., Huisman, J., Cohen-Kettenis, P.T., & Delemarre-van de Waal, HA (2008). An overview of studies on early development, cognition, and psychosocial well-being in children born after in vitro fertilization. *Journal of Developmental and Behavioral Pediatrics, 29 (3)*, p. 219-230.

Webbink, D., Posthuma, D., Boomsma, D.I., de Geus, E.J., & Visscher, P.M. (2008). Do twins have lower cognitive ability than singletons? *Intelligence, 36 (6)*, p. 539-547.

Wilson, C.L., Fisher, J.R., Hammarberg, K., Amor, D.J., & Halliday, J.L. (2011). Looking downstream: a review of the literature on physical and psychosocial health outcomes in adolescents and young adults who were conceived by ART. *Human Reproduction, 26 (5)*, p. 1209-1219.

Wright, V.C., Chang, J., Jeng, G., Macaluso, M., & Centers for Disease Control and Prevention (CDC) (2008). Assisted reproductive technology surveillance - United States, 2005. *MMWR Surveillance Summaries, 57 (5)*, p. 1-23.

3.

DO CHILDREN PERFORM AND BEHAVE BETTER AT SCHOOL WHEN TAUGHT BY SAME-GENDER TEACHERS?

Based on de Zeeuw, EL, van Beijsterveldt, CEM, Glasner, TJ, de Geus, EJC, Boomsma, DI (2014). Do children perform and behave better at school when taught by same-gender teachers? *Learning and Individual Differences*, *36*, p. 152-156.

An increase in the educational achievement of girls at the same time that the number of female teachers in primary school education is rising has led to the discussion whether boys are disadvantaged by the lack of male teachers. The Netherlands Twin Register identified a unique sample of 100 12-year-old monozygotic twin pairs discordant and 396 boy–girl twin pairs concordant for teachers' gender. School performance, as rated by the teacher, an educational achievement test score and teacher-rated ADHD behavior were similar for students with male or female teachers. In spite of the increase in the number of female teachers, boys still outperformed girls in arithmetic, while girls scored higher on language and reading. Boys demonstrated more ADHD behavior, but this was independent of teacher' gender. Therefore, increasing the number of male teachers in primary education may not be as effective to close a possible gender gap as suggested by some.

INTRODUCTION

In many Western countries the share of female teachers in the educational system is greater than ever. In the Netherlands, for example, the percentage of female teachers in primary education is over 80%, ranging from 40% in last grade to over 90% in first grade (Ministry of OCW, 2010). The share of female teachers in primary school will probably continue to rise since over 90% of teaching graduates is female (CBS Statistics Nederland, 2011). The majority of male and female teachers considers the feminization of primary education a problem for children (Sikkes, 2004). Female teachers are suggested to perceive the behavior of boys as more problematic and harm their social and emotional development and motivation to learn. Educational achievement of girls is increasing compared to boys at the same time that the share of women in primary education is rising. Reading and arithmetic levels are equal for 5-year old boys and girls attending kindergarten (Rathbun & West, 2004) From primary school onwards, girls receive higher grades for reading and language, and boys for mathematics and science (OECD, 2010). Furthermore, boys have to repeat a grade more often and more boys attend specialized education (Ministry of OCW, 2010). In college, the number of female students exceeds the number of males enrolled, except in the more technical fields (CBS Statistics Nederland, 2011). In many Western countries some people blame the feminization of primary education for the supposed gap in educational achievement between boys and girls as same-gender teachers are said to enhance educational achievement (Ailwood, 2003; Ammermuller & Dolton, 2006; Helbig, 2010). Policies to increase the number of male teachers and to promote single-gender education have both been proposed.

The suggestion that boys need male teachers to achieve their true learning potential has been around for some time. Yet, empirical findings of studies investigating the influence of the gender of the teacher on educational achievement are inconclusive. Some studies did not observe an influence of teacher's gender on educational achievement in primary or secondary school (Ehrenberg, Goldhaber & Brewer, 1995; Holmlund & Sund, 2008; Neugebauer, Helbig & Landmann, 2010; Sokal et al., 2007). On the other hand, a number of studies from the United States and other countries reported an enhancing influence of a same-gender teacher on school achievements (Klein, 2004) in mathematics (Ammermuller & Dolton, 2006; Dee, 2007; Helbig, 2010) and reading (Dee, 2007; Helbig, 2010). Sometimes students were rewarded higher grades (Ehrenberg, Goldhaber & Brewer, 1995) and more positive evaluations with regard to their educational achievement (Hopf & Hatzichristou, 1999) by a teacher of their own gender. The age of the children included in these studies varies widely and some assess children in primary school while others report on secondary school students. A number of studies compared grades while others used results on standardized tests. Confounding influences of student traits, amongst others, socioeconomic status, ethnicity, intelligence, and existing behavioral problems, could also be an explanation for the inconsistent results in the literature so far. Nonrandom placement with a male or female teacher with regard to these factors might confound results.

The enhancing influence of a same-gender teacher may be due to the fact that students identify more with a same-gender teacher and therefore work harder and behave better (Carrington, Tymms & Merrell, 2008). Alternatively, teachers may prefer or feel more competent with students whose gender they share, and encourage them more (Powell & Downey, 1997). Also, negative gender stereotypes can influence the way teachers perceive and interact with their students, and have a detrimental effect on student motivation (Steele, 1997). Some of these explanations suggest female teachers have a negative effect on the behavior of the boys in their classroom leading to underperformance (Carrington, Tymms & Merrell, 2008). Studies testing this hypothesis are rather scarce. One study in Dutch primary school students concluded that female teachers reported slightly more externalizing problem behavior for boys, not girls, compared to male teachers. However, parents of those twins did not rate the behavior of the children with a male and female teacher differently (Rietveld, van Beijsterveldt & Boomsma, 2010). A study from Greece reported that female teachers evaluated the behavior of their students more positively compared to male teachers (Hopf & Hatzichristou, 1999).

Our study adopts a unique design which makes it possible to minimize random error resulting from differences between children while also controlling for possible confounding by genetic influences. We employed two genetically sensitive designs, a discordant monozygotic (MZ) twin design and a concordant dizygotic of opposite sex (DOS) twin design. In MZ sample, twin pairs, who share nearly all their genes and always have the same gender, attended separate classes (or different schools). One twin had a male teacher and the co-twin was taught by a female teacher. In the DOS sample, twin pairs were included, who were both taught by the same or a different male or female teacher. Differences within the twin pairs may be ascribed to the influence of the gender of the teacher, since (part of their) genotype, family background, social economic status and multiple other characteristics of the twins are similar and controlled for.

METHODS

PARTICIPANTS

The Netherlands Twin Register (NTR), established around 1987 by the Department of Biological Psychology at the VU University Amsterdam, registers approximately 40% of all multiple births in the Netherlands (Boomsma et al., 2006; van Beijsterveldt et al., 2013). Parents of the twins gave permission to approach their primary school teachers. From birth cohort 1989 onwards teachers of the 12-year old twins were asked to report their own gender and since then data collection has yielded surveys on 6643 children. Data were excluded if a child had a disease or handicap that interfered severely with daily functioning (N=96), if the child attended education for children with special needs (N=231), if the survey was filled out by someone other than the regular teacher (N=60), if teacher familiarity with the student was below average (N=40) and if no information on the gender of the teacher was available (N=163), resulting in data for 6053 children from 2593 complete (data on both twins) and 867 incomplete twin pairs.

For the MZ sample, MZ twin pairs for whom a male teacher had filled out the survey of one of the twins and a female teacher had filled out the survey of the co-twin were selected. This smaller sample included 129 discordant MZ twin pairs. A short survey was sent to the parents of these MZ twin pairs to obtain additional information and check discordance at the time of the teacher survey. Parents were asked to report, for all grades, whether the twins attended the same or separate classes and the gender of the teacher(s) of their children. When parents did not return the survey they were contacted by phone. The information obtained from the parents revealed that some twin pairs, at the time of the teacher survey, were not in separate classes or one or both were taught by a male as well as a female teacher. For some twin pairs data could not be checked as parents were no longer willing to participate or a phone number was disconnected. Therefore, the final sample included 100 MZ twin pairs (39 male pairs; 61 female pairs) who were discordant for teacher's gender at the time of the teacher survey. In general, the decision to separate the twins was made by the parents in agreement with the school. None of the parents reported that a deliberate decision was made to place one of the children with the male and the other with the female teacher. Twin pairs included in this study were in separate classes during most of their primary school years. The total number of male teachers during primary school education was, on average, around two for the children who were taught by a male teacher at the time of the survey compared to one for the children with a female teacher. Zygosity status was based on DNA polymorphisms (N=26) or on a parental questionnaire with 10 items about resemblance in appearance (N=74). With this last method correct zygosity classification of MZ twins is estimated to be around 97% (Rietveld et al., 2000).

For the DOS sample, twin pairs for whom both surveys were filled out by either a male teacher or a female teacher were selected. The total sample included data from 1862 children belonging to a DOS twin pair. Surveys were excluded if a teacher had indicated that he or she did not teach a child at least 4 days per week and not all school domains (N=659). These exclusion criteria were applied to ensure that a child was not taught by both a male and a female teacher. This left surveys for 1203 children from 446 complete and 311 incomplete twin pairs. Surveys from twin pairs discordant for gender of their teacher were excluded. Resulting in data from 396 complete DOS twin pairs concordant for gender of their teacher, sharing either the same (N=167) or different (N=38) male or the same (N=158) or different (N=33) female teachers. Table 1 gives the background characteristics of the parents of the MZ and DOS twin pairs.

TABLE 1 Parental characteristics for the monozygotic (MZ) and dizygotic of opposite-sex (DOS) twin pairs

	MZ Twin Pairs		DOS Twin Pairs	
	N	%	N	%
Social Economic Status				
Lower Profession	11	11.0	61	15.4
Secondary Profession	40	40.0	168	42.4
Higher Profession	30	40.0	82	20.7
Scientific Profession	18	18.0	57	14.4
Missing	*1*	*1.0*	*28*	*7.1*
Maternal Age at Birth				
≤ 24 years	6	6.0	13	3.3
25-29 years	40	40.0	108	27.3
30-34 year	40	40.0	194	49.0
≥ 35 years	14	14.0	78	19.7
Missing	0	0.0	*3*	0.8
Paternal Age at Birth				
≤ 24 years	3	3.0	3	0.8
25-29 years	23	23.0	66	16.7
30-34 year	43	43.0	156	39.4
≥ 35 years	29	29.0	162	40.9
Missing	*2*	*2.0*	*9*	*2.3*
Education of the Mother				
Lower Education	2	2.0	17	4.3
Lower Middle Education	23	23.0	104	26.3
Higher Middle Education	33	33.0	151	38.1
Higher Education	41	41.0	95	24.0
Missing	*1*	*1.0*	*29*	*7.3*
Education of the Father				
Lower Education	5	5.0	27	6.8
Lower Middle Education	26	26.0	86	21.7
Higher Middle Education	27	27.0	129	32.6
Higher Education	40	40.0	126	31.8
Missing	*2*	*2.0*	*28*	*7.1*
Country of Birth of the Mother				
The Netherlands	97	97.0	373	94.2
other Western Country	0	0.0	11	2.8
non Western Country	2	2.0	3	0.8
Missing	*1*	*1.0*	*9*	*2.3*
Country of Birth of the Father				
The Netherlands	90	90.0	372	93.9
other Western Country	4	4.0	9	2.3
non Western Country	4	4.0	4	1.0
Missing	*2*	*2.0*	*11*	*2.8*

MEASUREMENTS

School performance was assessed by teacher ratings of several school domains, arithmetic, language and reading, with two versions of the teacher's survey. In the first version of the survey (birth cohorts 1989-1993), teachers could choose up to six domains and rate the proficiency of the students in these domains on a five-point scale from 1 (insufficient) to 5 (very good). In the second version (birth cohorts 1994-2000), teachers rated the proficiency of the students in four predefined school domains on the same five-point scale. Due to the free choice of school domains in the first version of the survey, the number of available teacher ratings differs across the school domains.

Educational achievement was assessed by a score on a national test of educational achievement which is administered at the end of the school term to grade 6 students (ages 11 to 13) at approximately 80% of all primary schools in the Netherlands (Cito, 2002). This test consists of multiple choice items in four different domains, namely arithmetic, language, study skills, and science and social studies. Together the scales of the national educational achievement test can be combined into a total score. Since administration of the science and social studies scale is not required, the science and social studies items are not used in the calculation of this standardized measure.

Behavioral problems were assessed with the ADHD Index scale of the short version of the Conners' Teacher Rating Scales - Revised (CTRS-R). The CTRS-R consists of 28 items scored on a 4 point scale from 0 (not true at all) to 3 (completely true) and includes 4 scales describing Oppositional Behavior, Inattention/Cognitive Problems, Hyperactivity and ADHD Index (Conners et al., 1998; Goyette, Conners & Ulrich, 1978). A sum score for the ADHD Index was only computed when a subject had two or less missing items on the scale. A missing item on a scale was imputed by the averaged item score within a scale of an individual child.

STATISTICAL ANALYSES

Data were analyzed using the Statistical Package for the Social Sciences 20.0 (IBM Corp, 2011). The main effects of teachers' gender, students' gender and the effect of their interaction on educational achievement, school performance and behavioral problems were examined with repeated measures general linear models. For the discordant MZ twin pairs the gender of the teacher was the within subjects factor and the gender of the student the between subjects factor while for the concordant DOS twin pairs the gender of the teacher was the between subjects factor and the gender of the student the within subjects factor. To correct for multiple testing a p-value of .01 (0.5/5 (total number of outcome measures)) was considered significant.

RESULTS

The estimates of means and standard errors for teacher-rated school performance, for the national educational achievement test, and for ADHD related behavior are given in Table 2 and displayed in Figure 1 for the 100 discordant MZ twin pairs and the 396 concordant DOS twin pairs. A summary of the results of the general linear model analyses is shown in Table 3. There were no significant main effects of teacher gender. In the MZ twin pairs sample, there was a main effect of student gender on ADHD ($p < .001$). Boys demonstrated more ADHD-related behavior than girls. In the DOS twin pairs sample, there was a main effect of student gender on teacher-rated school performance in arithmetic ($p < .001$), language ($p < .001$) and reading ($p < .001$) and on ADHD related behavior ($p < .001$). Boys received higher ratings for arithmetic and displayed more ADHD related behavior. Girls received higher grades for language and reading. There were neither significant interactions between student and teacher gender for teacher ratings for arithmetic, language and reading, nor for the educational achievement test and ADHD behavior.

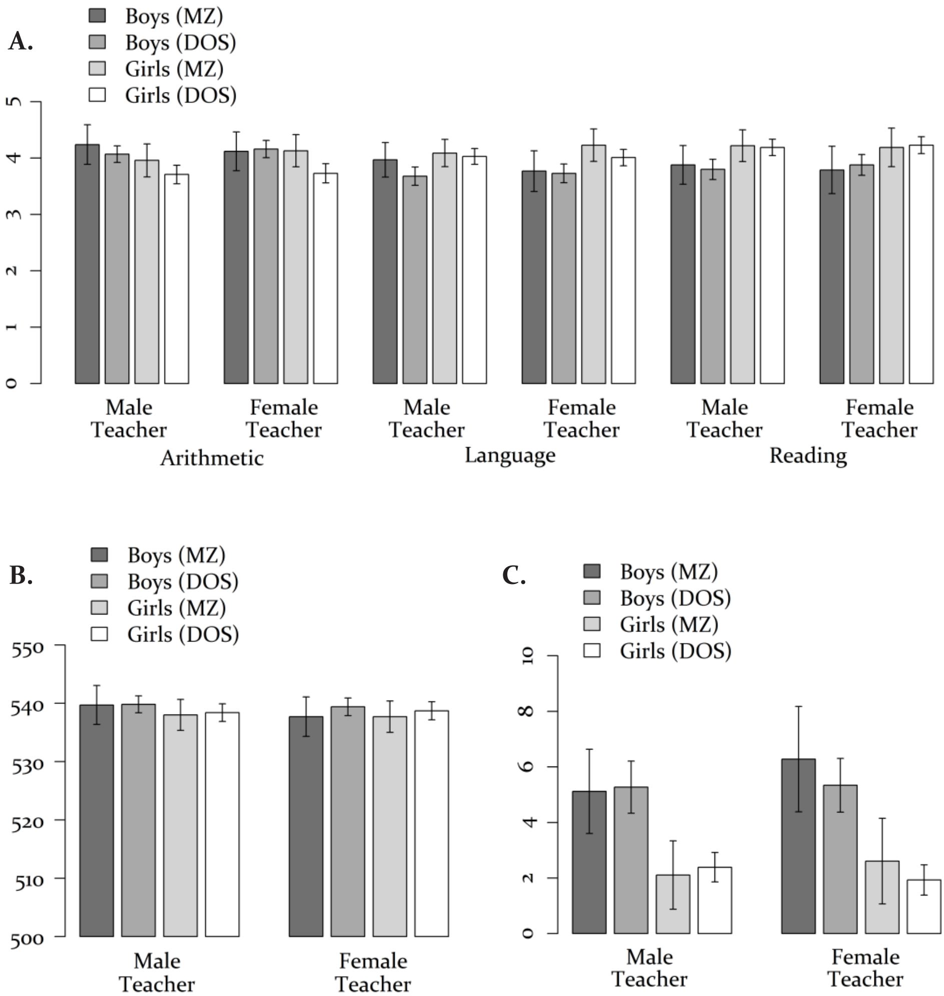
TABLE 2 Estimated means and standard errors from the general linear model analyses for the discordant monozygotic (MZ) twin pairs

	N Twin Pairs	Boys						Girls					
		Male Teacher			Female Teacher			Male Teacher			Female Teacher		
		N	Mean	SE	N	Mean	SE	N	Mean	SE	N	Mean	SE
MZ Twin Pairs													
School Performance													
Arithmetic	81	33	4.24	.179	33	4.12	.176	48	3.96	.149	48	4.13	.146
Language	77	30	3.97	.156	30	3.77	.184	47	4.09	.124	47	4.23	.147
Reading	60	24	3.88	.175	24	3.79	.214	36	4.22	.143	36	4.19	.175
Educational Achievement													
Total Score	67	26	539.7	1.70	26	537.7	1.73	41	538.0	1.36	41	537.7	1.37
Behavioral Problems													
ADHD	98	39	5.12	.772	39	6.28	.967	59	2.11	.628	59	2.61	.786
DOS Twin Pairs													
School Performance													
Arithmetic	374	193	4.07	.075	193	4.16	.078	181	3.71	.084	181	3.73	.087
Language	350	179	3.68	.083	179	3.73	.085	171	4.03	.072	171	4.01	.074
Reading	284	146	3.80	.092	146	3.88	.094	138	4.19	.074	138	4.23	.076
Educational Achievement													
Total Score	246	127	539.6	.74	127	539.4	.77	119	538.4	.77	119	538.7	.79
Behavioral Problems													
ADHD	391	201	5.27	.479	201	5.34	.493	190	2.39	.271	190	1.93	.279

TABLE 3 Results of the general linear model analyses for the monozygotic (MZ) and dizygotic of opposite-sex (DOS) twin pairs

		MZ Twin Pairs			DOS Twin Pairs		
		F (df)	**partial η^2**	**p**	**F (df)**	**partial η^2**	**p**
School Performance							
Arithmetic	Teacher	.033 (1,79)	.000	.856	.354 (1,372)	.001	.552
	Student	.520 (1,79)	.007	.473	36.198 (1,372)	.089	<.001
	Teacher x Student	1.331 (1,79)	.017	.252	.314 (1,372)	.001	.576
Language	Teacher	.052 (1,75)	.001	.820	.014 (1,348)	.000	.906
	Student	2.454 (1,75)	.032	.121	25.072 (1,348)	.067	<.001
	Teacher x Student	2.432 (1,75)	.031	.123	.379 (1,348)	.001	.538
Reading	Teacher	.220 (1,58)	.004	.641	.352 (1,282)	.001	.553
	Student	2.829 (1,58)	.047	.098	28.754 (1,282)	.093	<.001
	Teacher x Student	.055 (1,58)	.001	.815	.127 (1,282)	.000	.722
Educational Achievement Test							
Total Score	Teacher	2.156 (1,65)	.032	.147	.000 (1,244)	.000	.996
	Student	.172 (1,65)	.003	.680	4.224 (1,244)	.017	.041
	Teacher x Student	1.108 (1,65)	.017	.296	.164 (1,244)	.001	.686
Behavioral Problems							
ADHD	Teacher	1.615 (1,96)	.017	.207	.210 (1,389)	.001	.647
	Student	13.285 (1,96)	.122	<.001	75.194 (1,389)	.162	<.001
	Teacher x Student	.253 (1,96)	.003	.616	.516 (1,389)	.001	.473

FIGURE 1 Estimated means (95% CIs) of school performance assessed by teacher ratings **(A)**, educational achievement assessed with the national Dutch educational achievement test **(B)** and the assessment of ADHD behavior with the CTRS-R **(C)** for the monozygotic (MZ) and dizygotic of opposite-sex (DOS) twin pairs, by male and female teachers



DISCUSSION

Our study supplies some unique empirical data and sheds light on the debate concerning the feminization of primary school education and its influence on the educational achievement of children. It has been proposed that same-gender teachers enhance educational achievement and lessen behavioral problems. Male and female teachers do not rate the proficiency of their students in arithmetic, language and reading differently. They do not give higher ratings to children with whom they shared their gender. Looking at the score on a national educational achievement test, not rated by the teacher, there were also no differences between children with a male and a female teacher. Traditional gender differences were observed with boys outperforming girls on the numeracy domains while girls did better on the literacy domains and boys showed more ADHD-related behavior in the classroom. Teachers all throughout primary school contribute to a student's educational achievement and this study did not control for the gender of the teachers in the earlier grades. This means that we are mainly considering the short term effects, namely one school year, of a same-gender teacher. It could be that having a same-gender teacher all throughout primary school may have an effect on educational achievement and behavior.

A possible positive effect of a same-gender teacher may be due to female teachers negatively affecting the behavior of boys, leading to lower educational performance. This hypothesis is not supported by the findings from this study since there was no indication that male and female teachers rated the behavior of their students differently. Our findings are is in line with those from a Dutch longitudinal study among primary school children, which demonstrated that neither having a male teacher in the last grade of primary education nor the total number of male teachers affected educational achievement or social-emotional development of the students (Driessen, 2007).

Teaching quality varies and might have an effect on the educational achievement of students (Taylor et al., 2010). However, there appear to be no systematic differences in teaching methods between male and female teachers (Stone, 2010). Behaviors thought to be associated with masculinity as well as femininity are displayed by both male and female teachers (Skelton, 2003). Educational achievement could have been influenced by other factors, such as, classroom factors (e.g. classmates) and teacher characteristics (e.g. age), but only when they systematically differ between male and female teachers. For example, male teachers in primary school are, in general, older and therefore more experienced than their female colleagues which could perhaps influence the educational achievement of their students. The male teachers in our sample are in fact somewhat older than the female teachers, but additional analyses did not show an effect of age of the teacher on the different outcome measures.

The sample from which the MZ twin pairs were selected was rather large, but the number of MZ twin pairs discordant for teachers' gender was still small. As a consequence, the statistical power to detect the effects of interest was moderate in this group. The number of DOS twin pairs who were concordant for teachers' gender was larger, and thus had greater statistical power. A post-hoc power analysis in G*Power (Faul et al., 2007) revealed that, for the MZ twin pairs, the sample had enough power to detect an interaction between gender of the teacher and student of medium effect size (Cohen's $d = .5$) (Cohen, 1988) at a significance threshold of $p < .01$. The power for a small effect size (Cohen's $d = .2$) was low, ranging from .17 to .38 for the various outcome measures. The power of the DOS twin pairs sample to find small effect sizes was larger and ranged from .63 to .79. Results in both groups showed no effect of a same-gender teacher. Therefore, increasing the number of male teachers in primary education or implementing single-gender education may not be as effective to close a possible gender gap between the educational achievement of boys and girls as suggested by some.

REFERENCES

Ailwood, J. (2003). A national approach to gender equity policy in Australia: Another ending, another opening? *International Journal of Inclusive Education, 7 (1)*, p. 19-32.

Ammermuller, A. & Dolton, P.J. (2006). Pupil-teacher gender interaction effects on scholastic outcomes in England and the USA. ZEW Discussion Papers.

Boomsma, D.I., de Geus, E.J., Vink, J.M., Stubbe, J.H., Distel, M.A., Hottenga, J.J., Posthuma, D., van Beijsterveldt, T.C., Hudziak, J.J., Bartels, M., & Willemsen, G. (2006). Netherlands Twin Register: from twins to twin families. *Twin Resarch and Human Genetics, 9 (6)*, p. 849-857.

Carrington, B., Tymms, P., & Merrell, C. (2008). Role models, school improvement and the gender gap: Do men bring out the best in boys and women the best in girls? *British Educational Research Journal, 34 (3)*, p. 315-327.

CBS Statistics Nederland (2011). Hoger beroepsonderwijs en wetenschappelijk onderwijs: Geslaagden.

Cito (2002). Eindtoets Basisonderwijs. Arnhem, The Neterlands: Cito.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. New Jersey: Lawrence Erlbaum Associates.

Conners, C.K., Sitarenios, G., Parker, J.D., & Epstein, J.N. (1998). Revision and restandardization of the Conners Teacher Rating Scale (CTRS-R): factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology, 26 (4)*, p. 279-291.

Dee, T.S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources, 42 (3)*, p. 528-554.

Driessen, G. (2007). The feminization of primary education: effects of teachers' sex on pupil achievement, attitudes and behaviour. *International Review of Education, 53 (2)*, p. 183-203.

Ehrenberg, R.G., Goldhaber, D.D., & Brewer, D.J. (1995). Do teachers' race, gender, and ethnicity matter? Evidence from NELS88. *Industrial and Labor Relations Review, 48 (547)*, p. 561.

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39 (2)*, p. 175-191.

Goyette, C.H., Conners, C.K., & Ulrich, R.F. (1978). Normative data on revised Conners parent and teacher rating scales. *Journal of Abnormal Child Psychology, 6 (2)*, p. 221-236.

Helbig, M. (2010). Sind Lehrerinnen für den geringeren Schulerfolg von Jungen verantwortlich? *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 62 (1)*, p. 93-111.

Holmlund, H. & Sund, K. (2008). Is the gender gap in school performance affected by the sex of the teacher? *Labour Economics, 15 (1)*, p. 37-53.

Hopf, D. & Hatzichristou, C. (1999). Teacher gender related influences in Greek schools. *British Journal of Educational Psychology, 69 (1)*, p. 1-18.

IBM Corp (2011). SPSS Statistics for Windows Version 20.0.

Klein, J. (2004). Who is most responsible for gender differences in scholastic achievements: pupils or teachers? *Educational Research, 46 (2)*, p. 183-193.

Ministry of OCW (2010). Key figures 2005-2009 Education, culture and science. Den Haag, The Netherlands: Ministery of Education, Culture and Science.

Neugebauer, M., Helbig, M., & Landmann, A. (2010). Unmasking the myth of the same-sex teacher advantage. *European Sociological Review, 27 (5)*, p. 669-689.

OECD (2010). PISA 2009 Results: What students know and can do - Student performance in reading, mathematics and science (Volume I).

Powell, B. & Downey, D.B. (1997). Living in single-parent households: An investigation of the same-sex hypothesis. *American Sociological Review,* p. 521-539.

Rathbun, A. & West, J. (2004). From kindergarten through third grade: children's beginning school experiences (NCES 2004-007). Washington, DC: US Department of Education.

Rietveld, M.J.H., van Beijsterveldt, C.E.M., & Boomsma, D.I. (2010). Sekse van de leerkracht en probleemgedrag bij leerlingen. *Pedagogische Studiën, 88* p. 59-72.

Rietveld, M.J.H., van Der Valk, J.C., Bongers, I.L., Stroet, T.M., Slagboom, P.E., & Boomsma, D.I. (2000). Zygosity diagnosis in young twins by parental report. *Twin research, 3 (3)*, p. 134-141.

Sikkes, R. (2004). Money, status, and boys' tender souls. A survey on the feminization of education. *Het Onderwijsblad, 8* p. 16-23.

Skelton, C. (2003). Male primary teachers and perceptions of masculinity. *Educational Review, 55 (2)*, p. 195-209.

Sokal, L., Katz, H., Chaszewski, L., & Wojcik, C. (2007). Good-bye, Mr. Chips: Male teacher shortages and boys' reading achievement. *Sex Roles, 56 (9-10)*, p. 651-659.

Steele, C.M. (1997). A threat in the air: how stereotypes shape intellectual identity and performance. *American psychologist, 52 (6)*, p. 613-629.

Stone, K. (2010). Dealing with gender disparity in education: More men and only boys.

Taylor, J., Roehrig, A.D., Soden, H.B., Connor, C.M., & Schatschneider, C. (2010). Teacher quality moderates the genetic effects on early reading. *Science, 328 (5977)*, p. 512-514.

van Beijsterveldt, C.E., Groen-Blokhuis, M., Hottenga, J.J., Franic, S., Hudziak, J.J., Lamb, D., Huppertz, C., de Zeeuw, E., Nivard, M., Schutte, N., Swagerman, S., Glasner, T., van, F.M., Brouwer, C., Stroet, T., Nowotny, D., Ehli, E.A., Davies, G.E., Scheet, P., Orlebeke, J.F., Kan, K.J., Smit, D., Dolan, C.V., Middeldorp, C.M., de Geus, E.J., Bartels, M., & Boomsma, D.I. (2013). The Young Netherlands Twin Register (YNTR): longitudinal twin and family studies in over 70,000 children. *Twin Research and Human Genetics, 16 (1)*, p. 252-267.

Part II

GENETIC FACTORS

4.

ARITHMETIC, READING AND WRITING PERFORMANCE HAS A STRONG GENETIC COMPONENT: A STUDY IN PRIMARY SCHOOL CHILDREN

Based on de Zeeuw, E.L., van Beijsterveldt, C.E.M., Glasner, T.J., de Geus, E.J.C., & Boomsma, D.I. (2015). Arithmetic, reading and writing performance has a strong genetic component: A study in primary school children. In Preparation.

Most research on educational achievement in children has focused on environmental factors and on differences between groups. However, even children taught by the same teacher differ greatly in their performance at school. Genetic research can address the extent to which these individual differences can be explained by genes or the environment. The current study aims to identify the impact of genes on variation in educational achievement in a large cohort of Dutch children (6 to 12 years). The Netherlands Twin Register has collected data on pupil monitoring tests used in all grades of primary school to measure a child's educational achievement in four educational domains, i.e. arithmetic, reading, reading comprehension and spelling (1058 MZ and 1734 DZ twin pairs) and on an educational achievement test administered in the last grade (2451 MZ and 4569 DZ twin pairs). Genes were the most important cause of differences between children in arithmetic (60-74%), reading (73-82%) and reading comprehension (54-63%) across all grades. The common environment, i.e. socioeconomic status and the school environment had, in general, only a minor influence on educational achievement. In contrast, heritability of spelling was small in the first grade (33%), compared to later ages (58-70%), with a larger influence of the common environment (28%). Heritability of the educational achievement test was also large (74%) with a small influence of the common environment (8%). The heritability of children's educational achievement in The Netherlands is surprisingly comparable to other countries despite major differences in educational systems and teaching methods.

INTRODUCTION

Children differ in their ability to learn the subject material that is taught at school; some master basic skills, such as reading and arithmetic, and pick up knowledge about science, history and biology much faster than their peers. Low educational achievement is associated with continued low achievement, school dropout and delinquency (Moilanen, Shaw & Maxwell, 2010). General cognitive ability is the most important predictor of educational achievement (Deary et al., 2007) and explains about half of the variation (Frey & Detterman, 2004). Most research towards educational achievement of children has focused on environmental factors, such as socioeconomic status (SES) of the parents and school characteristics, and on differences between groups of children, for example boys and girls (OECD, 2010). However, even children attending the same school and taught by the same teacher differ greatly in their performance at school. It may be less relevant to look at group differences when differences within a group are much larger. Causes for individual differences do not necessarily have to be the same as for average differences between groups. The main reason for mean differences in educational achievement between boys and girls might be environmental whereas the cause for differences in performance between individual children may be largely genetic in nature.

Genetic research can address questions about the causes of individual differences among children and disentangle the extent to which these differences in educational achievement between children are explained by their genes or by the environment (Boomsma, 2013; Plomin et al., 2008). One of the most often used designs in behavior genetics is the twin study, which is based on the difference in genetic relatedness between identical or monozygotic (MZ) and fraternal or dizygotic (DZ) twin pairs. MZ twins develop from the same fertilized zygote and are genetically (nearly) identical while DZ twins develop from separate eggs and share, like non-twin siblings, approximately 50 per cent of their segregating genes. When MZ twins are more similar for a certain phenotype than DZ twins this constitutes evidence for the influence of genetic effects. The environment can be distinguished in the common environment, such as SES of the parents, which is shared between MZ as well as DZ twins, and make them more similar, and the unique environment (including measurement error), which is not shared between twins. When DZ twins are similar to the same extent as MZ twins the common environment has an influence on a phenotype. When MZ twins are dissimilar this indicates that unique environmental effects contribute to individual differences in a phenotype.

Numerous studies have demonstrated that genetic effects have a substantial influence on differences between children in general cognitive ability. It is well established that the heritability of general cognitive ability increases from approximately 40 to 65 per cent from childhood into adulthood (Haworth et al., 2010). Children, when they grow up, can more and more select their own environments based on their genetic make-up and this may be one explanation for this increase (Molenaar et al., 2013). General cognitive ability is often seen as an aptitude while reading, mathematics and spelling are taught at school and perceived as the outcome of education. Hence, it seems reasonable to expect that heritability of educational achievement is lower than the heritability of general cognitive ability. However, a recent study has shown that the opposite was true for primary school children in the United Kingdom (UK). Literacy and numeracy were significantly more heritable than general cognitive ability at ages 7 and 9, but no longer at age 12 (Kovas et al., 2013). The authors propose that the equal opportunities in the relatively homogenous education environment provided in Western societies acts to reduce environmental variation, making differences in educational achievement between children to a greater extent due to genetic differences.

Twin studies have mainly focused on the educational domain reading and, more recently, mathematics, while less is known about the heritability of other educational domains, such as, spelling, reading comprehension and science. Most studies used teacher assessments or tests that had been administered by the researchers through the internet, telephone or during a home-visit while only some used standardized tests administered at school. Heritability of educational achievement in reading is moderate to high with modest common environmental effects (Byrne et al., 2013; Harlaar et al., 2012; Hart et al., 2013; Kovas et al., 2013) and the same is true for mathematics (Harlaar et al., 2012; Kovas et al., 2013). The twin studies towards educational achievement primarily included English speaking children from the USA and the UK. Studies from other countries with different educational systems and languages are scarce (Byrne et al., 2009; Chow et al., 2011). Previous research has established that languages differ in the complexity of their orthography and it is demonstrated that learning to read is more difficult in English than in other languages (Caravolas et al., 2013; Seymour, Aro & Erskine, 2003). The relationship between the printed words and phonemes in spoken words is least consistent in English. On the other hand, the number naming system is more consistent in English than it is in other languages which could positively influence learning to count and calculate compared to other countries (Göbel et al., 2014).

The question is whether the same pattern of estimates of the relative contribution of genetic and environmental effects to the variation in educational achievement exists in the Netherlands. (Calvin et al., 2012) found genetic effects to be an important cause of variation in educational achievement in the educational domains language (43-74%) and arithmetic (36-73%) at age 8, 10 and 12 in Dutch primary school children. However, they used a population cohort without information on zygosity and estimated the resemblance between monozygotic and dizygotic from the proportion of same-sex and opposite-sex twin pairs, but this method is much less powerful than a design in which zygosity is known.

In the Dutch educational system, the majority of primary schools use a pupil monitoring system that includes standardized tests assessing educational achievement (Cito, 2014a; Vlug, 1997). Tests are available for all grades and all important educational domains. The tests are independent of teaching methods and can be used to monitor a child's educational development in comparison to peers and across grades and educational domains. Tests measuring arithmetic, reading, reading comprehension and spelling are, according to an inventory amongst teachers, the most informative with regard to the educational development of children (Polderman et al., 2011). A standardized educational achievement test is available for the last grade, measuring what a child has learned during all primary school years (Cito, 2002). Together, these data provide a unique opportunity to give an overview of the heritability of educational achievement in different educational domains across primary school grades. Very few studies examined differences between boys and girls in the heritability of educational achievement, probably due to the small sample sizes. Quantitative gender differences are present if one gender is affected to a greater extent by the same genetic or environmental effects. Qualitative gender differences exist when different genetic or environmental effects have an influence on boys and girls. The aim of the current study is to identify the impact of genes and the environment on arithmetic, reading, reading comprehension and spelling in primary school in a large cohort of Dutch twins and to explore possible gender differences.

METHODS

PARTICIPANTS

The Netherlands Twin Register (NTR), established around 1987 by the Department of Biological Psychology at the VU University Amsterdam, registers approximately 40 per cent of all multiple births in the Netherlands. The parents of the twins receive a survey about the development of their children every two years up until the twins are 12 years old. Details about selection and response rates are described elsewhere (Bartels et al., 2007; Boomsma et al., 2002; Boomsma et al., 2006; van Beijsterveldt et al., 2013). Since 1999, at age 7, 9 and 12, when the twins attend primary school, parents are asked for their consent to approach the teacher(s) of their children. The survey sent to the primary school teachers consists of questions about behavioral and emotional problems, functioning at school and educational achievement. In addition, teachers are requested to provide information on results of the pupil monitoring tests. Results on a standardized educational achievement test, which is administered in the last grade of primary school, were also obtained from the teachers (Cito, 2002). Later, because results become available near the end of the last school year, parents were asked to report the scores of their children on this test.

Data on one of the pupil monitoring tests and/or the educational achievement test were available for 16234 children. We excluded children who had a disease or handicap that interfered severely with daily functioning (N=90) or attended specialized education (N=79), in the Dutch education system special schools are available for children who need extra care due to learning problems, physical and/or mental disabilities or a behavioral disorder, resulting in data for 7228 complete and 1609 incomplete twin pairs. One of the main reasons for incomplete data is that twins attend different classes or schools and only one of the teachers returned the survey.

The study included data of 2818 twin pairs of opposite sex. For same-sex twin pairs, the determination of zygosity status was based on blood or DNA polymorphisms (N=1363) or on parental report of items on resemblance in appearance and confusion of the twins by parents and others (N=4586). This last method can establish zygosity with an accuracy of approximately 93 per cent (Rietveld et al., 2000). Twin pairs for which zygosity was unavailable were excluded from the analyses (N=70). Data on the educational achievement test were available for the majority of these twin pairs (1113 MZm, 1132 DZm, 1338 MZf, 1149 DZf and 2288 DOS) and results for at least one of the pupil monitoring tests were available for approximately one third of the twin pairs (504 MZm, 465 DZm, 554 MZf, 428 DZf and 841 DOS).

MEASUREMENTS

The pupil monitoring system consists of tests to assess the educational achievement of a child in multiple educational domains (Cito, 2014a). The number of correct responses is converted in an ability score and these ability scores can be compared between grades to monitor a child's development in comparison to peers and over time. Each test score is converted into an ability score with a measurement technique on the basis of item-response theory to ensure that the development in ability scores is on a single scale (Vlug, 1997).

The arithmetic test (grade 1 to 6) consists of a part in which children have to solve simple math problems within a short time period and a part with more complex math problems without a time limit. The test assesses general knowledge of mathematics and arithmetic and comprises written computational problems of addition, subtraction, multiplication and division and problems on the notion of measurements, time and money, and knowledge about fractions, ratios and percentages.

The reading test (grade 1 to 6) measures word decoding skills by counting the total number of individual words a child can correctly read aloud in 1 minute. The test consists of three levels of increasing difficulty and complexity. The first level includes words that are pronounced exactly as they are spelled, the second level includes also other monosyllabic words and the third level includes two or more syllabic words. This study uses the most difficult level of the test which is almost never administered in the first grade.

The reading comprehension test (grade 3 to 6) includes a large variety of different text types and genres with two different types of multiple choice questions. The test consists of a part in which a child has to read a number of short texts and answer questions related to the text and a part with parts of the text left blank that need to be filled out. The test tries to assess different components of reading processing by questions regarding both the facts and events described in the texts as well as by questions about the purpose of the writer and the intended readership of the texts.

The spelling test (grade 1 to 6) measures both active, writing down the words, and passive, recognizing spelling errors, spelling. Active spelling is measured with a dictation by the teacher were a sentence is read aloud and a child has to write down a specific word from this sentence. Passive spelling is measured with multiple choice questions were a student has to choose the sentence in which the bolded word is spelled incorrectly.

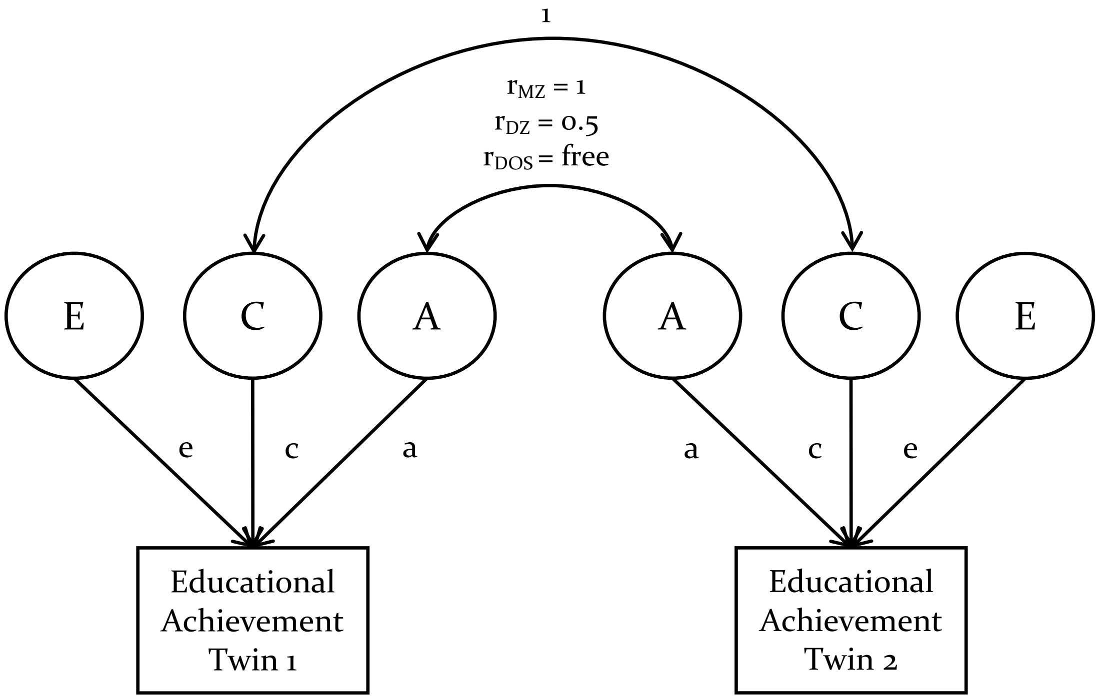
The educational achievement test measures what a child has learned during all primary school grades (Cito, 2002). The results of this test are often used, besides the advice of the teacher, to determine the level of secondary education suitable for a child. The test consists of multiple choice items in four different educational domains, namely Arithmetic, Language, Study Skills and Science and Social Studies. All scores on the scales are standardized to percentile scores to correct for differences in the number of items across the years. The first three test scales are combined into a Total Score, which is standardized on a scale from 500 and 550. The Arithmetic scale includes items on numbers and operations, ratios, fractions and percentages, and measurements, geometry, time and money. The Language scale includes items on writing, spelling, reading comprehension and vocabulary. The Study Skills scale includes items on handling of study texts, handling of information, reading diagrams, tables and graphs and map reading. The Science and Social Studies scale includes items on geography, history and biology.

STATISTICAL ANALYSES

Models were fitted to the raw data in the R (R Core Team, 2014) package OpenMx Version 3.0.3 (Boker S.M. et al., 2012; Boker S.M. et al., 2011) with maximum likelihood estimation. The correlations between the MZ and DZ twins were estimated, separately for each gender, to evaluate the relative influence of genetic and environmental effects on educational achievement. A model that freely estimated all parameters, i.e. means, variances and covariances, separately for the different zygosity-by-gender groups, was fitted to the data (saturated model). Norms and questions have been updated regularly for the tests for arithmetic, reading comprehension and spelling, resulting in different means and variances across versions. To correct for these differences, means and variances were estimated separately across different versions. Mean and variance differences between boys and girls were tested in the saturated model (Purcell, 2002).

To gain further insight into the causes of individual differences in educational achievement of children in primary school, univariate genetic models were fitted to the data for each educational domain and grade. The variation in educational achievement was decomposed into variance due to additive genetic effects (A), to common environmental effects (C) and to unique environmental effects (E) (Posthuma et al., 2003). Additive genetic effects are the sum of the effects of all genetic variants influencing educational achievement. Common environmental variance results from environmental effects that are shared by both members of a twin pair. Unique environmental variance results from environmental effects that are not shared by a twin pair. The variance components A, C and E were estimated separately for boys and girls. The variance components are expected to correlate differently for MZ and DZ twin pairs due to the difference in genetic resemblance (Figure 1). Since MZ twin pairs share (nearly) all their genes the correlation between the genetic effects of MZ twin pairs is fixed to 1.0. DZ twin pairs share approximately 50 per cent of their segregating genes and therefore the correlation between the genetic effects of DZ twin pairs is fixed to 0.5. The correlation between the genetic effects is estimated freely for DOS twin pairs as different genetic effects could have an influence on educational achievement in boys and girls. For both MZ, DZ and DOS twin pairs the correlation between the common environmental effects is fixed to 1.0.

FIGURE 1 Path diagram representing the twin model



A = additive genetic effects; C = common environmental effects; E = unique environmental effects; rMZ = correlation between MZ twins; rDZ = correlation between DZ twins; rDOS = correlation between DOS twins; a = coefficient representing the path loading for the additive genetic effects; c = coefficient representing the path loading for the common environmental effects; e = coefficient representing the path loading for the unique environmental effects
Note: the coefficients of the path loadings are estimated separately for boys and girls

In the subsequent models, the influence of the gender of the student on the variance components was tested in two ways. First, we tested whether the same genetic effects are important in boys and girls (qualitative gender differences) by fixing the correlation between the genetic effects of DOS twin pairs to be equal to the correlation for DZ twin pairs. Qualitative gender differences will result in a lower genetic correlation between DOS twin pairs. Second, we tested whether the genetic effects had an influence to the same extent in boys and girls (quantitative gender differences) by fitting a model, which incorporates total variance differences, but does not allow the relative contribution of the variance components to be different between boys and girls. Quantitative gender differences will result in unequal variance components between boys and girls. Finally, the significance of the common environmental effects was tested by dropping them from the model. The difference in model fit between the nested models was assessed with a log-likelihood ratio test (LRT) which calculates the difference in -2log-likelihood (-2LL) and evaluates this χ^2-statistic with the difference in the number of estimated parameters between the models as degrees of freedom. A p-value smaller than 0.01 was considered significant. Constraints were kept, when a more restrictive model did not significantly decrease the goodness of fit, as a more parsimonious model is preferred.

RESULTS

Table 1 gives the means and standard deviations of scores on the pupil monitoring tests across all grades and the educational achievement test, for boys and girls separately and for the old and the new version of the tests. There were significant gender differences for arithmetic and reading comprehension in most grades. Boys were better at arithmetic and girls performed better on the reading comprehension tests. Gender differences were also present for all scales of the educational achievement test. Boys scored higher on arithmetic, study skills and science and social studies while girls obtained better results for language.

Twin correlations were estimated in the five zygosity-by-gender groups and could be equated between the different versions of the tests (Table 2). All MZ correlations were higher than DZ correlations, suggesting additive genetic effects. Sometimes DZ correlations were larger than half the MZ correlations, suggesting common environmental effects. The genetic model fitting results with the standardized estimates and their 95% confidence intervals are reported for arithmetic, reading, reading comprehension and spelling across grades and an educational achievement test administered in the last grade of primary school (Table S1). The full model showed small differences between boys and girls in the heritability estimates, but these were not significant for all but one test, spelling in grade 5. The relative contribution of the additive genetic, common environmental and unique environmental effects are displayed for the models estimating all variance components equal for boys and girls while allowing total variance differences (Figure 2).

TABLE 1 Means and standard deviations of the educational achievement test scores

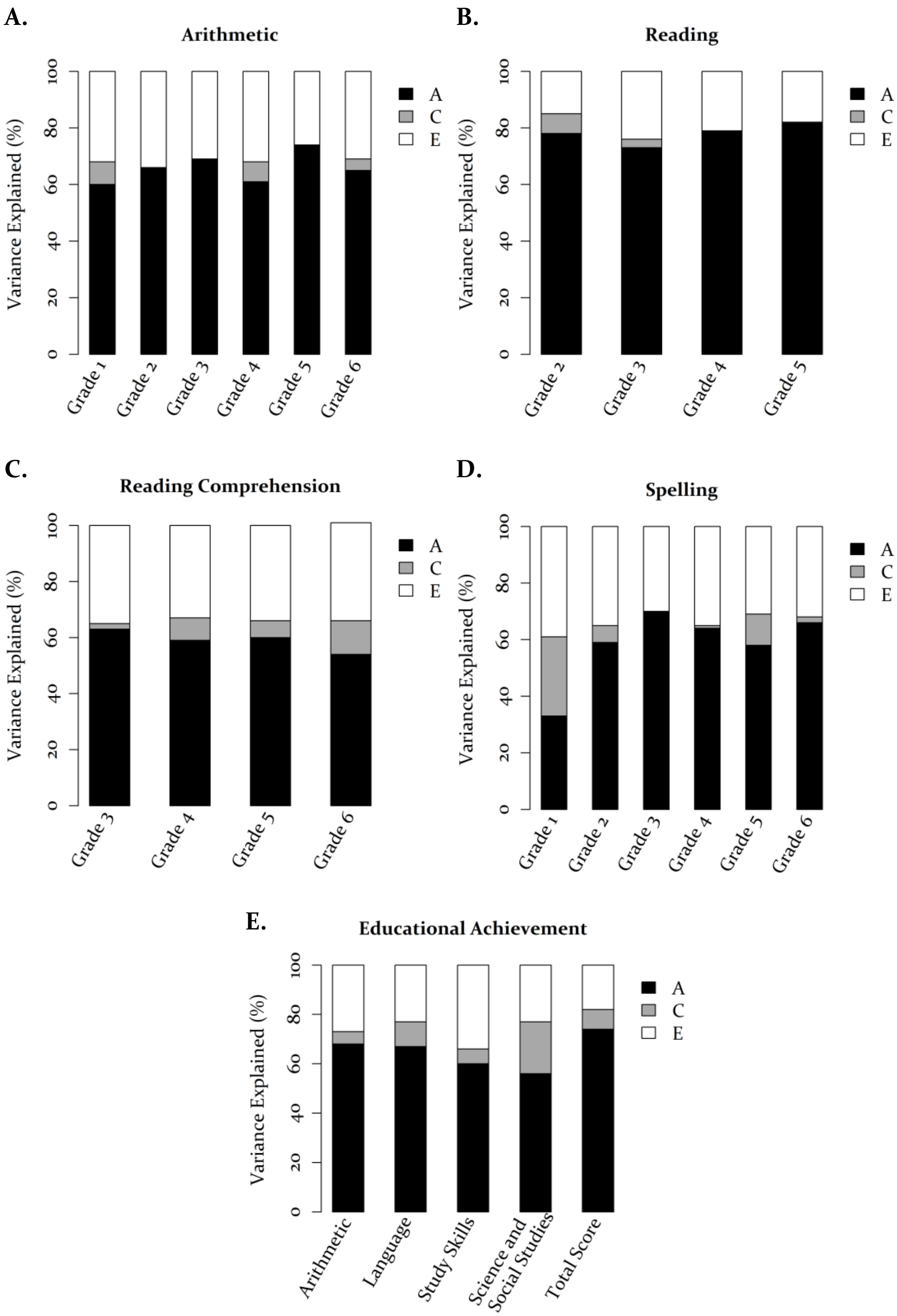
	Old Version				New Version			
	Boys		Girls		Boys		Girls	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Arithmetic								
Grade 1	429	53.9 (12.0)	418	52.9 (12.6)	1097	37.6 (15.8)	1093	34.4 (14.5)
Grade 2	491	74.7 (11.5)	535	69.7 (11.2)	857	57.6 (14.5)	854	51.5 (13.3)
Grade 3	536	91.0 (11.0)	555	86.0 (11.2)	800	79.0 (12.0)	805	73.0 (14.0)
Grade 4	560	100.1 (9.1)	593	96.3 (9.5)	439	91.4 (11.4)	442	86.0 (11.9)
Grade 5	515	111.8 (8.0)	530	108.1 (8.5)	434	104.4 (10.4)	410	100.6 (11.2)
Grade 6	355	123.4 (7.8)	399	120.2 (7.8)	116	118.2 (11.7)	111	112.8 (10.0)
Reading								
Grade 2	852	41.9 (18.6)	883	42.7 (19.2)				
Grade 3	900	63.2 (18.1)	910	63.8 (18.8)				
Grade 4	683	74.6 (17.3)	693	74.8 (16.8)				
Grade 5	598	84.1 (16.4)	605	86.0 (15.3)				
Reading Comprehension								
Grade 3	693	30.7 (14.1)	717	33.3 (13.3)	704	29.5 (14.1)	690	31.7 (13.4)
Grade 4	751	41.9 (14.2)	785	42.0 (13.7)	325	36.6 (13.7)	348	36.5 (12.4)
Grade 5	721	50.5 (14.9)	728	52.4 (15.0)	327	48.9 (13.7)	346	49.7 (15.5)
Grade 6	554	62.1 (15.0)	586	64.7 (15.1)	164	64.3 (16.8)	186	64.5 (17.8)
Spelling								
Grade 1	693	109.3 (8.9)	634	110.6 (8.6)	889	108.7 (7.3)	961	109.8 (7.0)
Grade 2	764	119.1 (7.5)	750	120.3 (7.2)	690	120.2 (6.9)	699	121.5 (6.7)
Grade 3	738	127.3 (7.0)	751	128.1 (6.5)	674	127.9 (7.1)	672	129.2 (7.1)
Grade 4	775	136.2 (7.2)	813	136.9 (7.0)	304	134.3 (7.5)	316	135.5 (7.2)
Grade 5	723	142.3 (6.5)	761	143.2 (6.5)	302	139.6 (6.5)	313	140.5 (5.8)
Grade 6	561	150.0 (5.9)	587	150.1 (6.0)	143	145.5 (9.5)	155	145.7 (8.1)
Educational Achievement								
Arithmetic	4330	65.3 (26.1)	4678	53.2 (28.2)				
Language	4327	57.8 (27.5)	4681	60.7 (26.9)				
Study Skills	4322	62.5 (27.1)	4677	59.7 (28.0)				
Science and Social Studies	3829	66.3 (25.9)	4133	51.6 (27.3)				
Total Score	6137	538.4 (8.4)	6797	537.2 (8.7)				

TABLE 2 Twin correlations for arithmetic, reading, reading comprehension, spelling and educational achievement

	N	MZm	N	DZm	N	MZf	N	DZf	N	DOS
Arithmetic										
Grade 1	341	.664	292	.449	348	.721	271	.326	568	.386
Grade 2	302	.667	271	.428	326	.660	253	.320	510	.318
Grade 3	301	.659	269	.397	319	.733	248	.272	509	.332
Grade 4	224	.704	201	.447	243	.690	194	.405	378	.296
Grade 5	212	.727	190	.320	215	.696	176	.396	356	.399
Grade 6	119	.620	93	.557	124	.682	96	.458	179	.326
Reading										
Grade 2	194	.822	175	.434	215	.864	161	.459	328	.488
Grade 3	199	.733	182	.382	215	.748	174	.475	369	.386
Grade 4	147	.774	150	.520	156	.789	136	.298	276	.327
Grade 5	145	.849	125	.680	144	.805	118	.371	224	.328
Reading Comprehension										
Grade 3	305	.665	285	.407	327	.665	261	.300	535	.339
Grade 4	232	.710	219	.513	259	.651	215	.309	411	.333
Grade 5	228	.649	215	.417	253	.717	202	.287	390	.375
Grade 6	167	.849	147	.545	179	.784	147	.432	275	.258
Spelling										
Grade 1	344	.623	316	.399	365	.567	281	.399	594	.524
Grade 2	320	.648	283	.405	345	.669	262	.336	544	.356
Grade 3	311	.689	285	.329	338	.709	255	.272	533	.294
Grade 4	234	.699	225	.446	263	.605	215	.271	412	.303
Grade 5	224	.704	211	.306	255	.675	199	.596	393	.334
Grade 6	166	.721	141	.483	181	.687	144	.374	259	.201
Educational Achievement Test										
Arithmetic	757	.700	787	.374	907	.757	765	.408	1618	.369
Language	757	.765	787	.457	908	.781	765	.423	1617	.449
Info Skills	755	.633	786	.375	908	.697	765	.350	1618	.359
Science and Social Studies	668	.745	695	.515	802	.785	677	.455	1448	.472
Total Score	1112	.804	1129	.468	1337	.830	1149	.432	2280	.437

MZm = monozygotic male twin pairs; MZf = monozygotic female twin pairs; DZm = dizygotic male twin pairs; DZf = dizygotic female twin pairs; DOS = dizygotic of opposite sex twin pairs; N= number of (in)complete twin pairs

FIGURE 2 The relative contribution of the genetic, common environmental and unique environmental effects for arithmetic (A), reading (B), reading comprehension (C), spelling (D) and educational achievement (E)



Genetic effects were the most important contributor to individual differences in educational achievement in arithmetic (60-74%), reading (72-82%) and reading comprehension (54-63%) and for most grades in spelling (33-70%). Common environmental effects had a negligible influence on arithmetic (0-8%), reading (0-7%) and reading comprehension (2-12%) and a slightly larger influence on spelling (0-29%). Unique environmental effects explained the remaining variance in arithmetic (26-34%), reading (11-29%), reading comprehension (32-35%) and spelling (30-39%). Genes were also the largest contributor to the variation in the educational achievement test (74%). The heritability differed somewhat between the educational domains measured with this test, i.e. arithmetic (68%), language (67%), study skills (60%) and science and social studies (56%). The common environmental effects were also small for the total score (8%), arithmetic (5%), language (10%), study skills (6%) and science and social studies (21%). Unique environmental effects explained the remaining variance (18-34%).

DISCUSSION

The current study presents the heritability of educational achievement in several educational domains across primary school grades 1 to 6, corresponding to ages 6 to 12, in a large, representative cohort from The Netherlands. The influence of the genetic and environmental effects was systematically examined for the educational domains arithmetic, reading, reading comprehension and spelling. The extent to which genes influenced differences in educational achievement was large and relatively stable across all grades for arithmetic (60-74%), reading (72-82%) and reading comprehension (54-63%). Similar heritability estimates were obtained at all ages despite large differences in content across grades. In contrast, the heritability of spelling was smaller in the first grade (33%) compared to later ages (58-70%).

Heritability of the educational achievement test in the last grade was high, higher than estimated in a partly overlapping, but considerably smaller, sample (74 vs 57%) while the influence of the common environment was slightly lower (8 vs 27%) (Bartels et al., 2002). There was a noteworthy difference between the heritability of the specific domains included in the educational achievement test. Science and social studies, i.e. geography, history and biology, showed somewhat lower heritability estimates and a larger influence of the common environment compared to arithmetic and language. This is also observed in earlier research towards science performance which too is not one of the core educational domains, e.g. writing, reading and mathematics (Haworth et al., 2008).

Traditional mean gender differences in educational achievement were observed with boys scoring better on numeracy and girls performing better at some of the literacy subjects (Cito, 2014b; OECD, 2010). There was no consistent indication for the presence of quantitative gender differences, meaning that the extent to which genes and the environment influence educational achievement is similar across gender. Qualitative gender differences were also not present which means that the genes that have an influence on educational achievement are the same for boys and girls. The absence of gender differences in the heritability of educational achievement is in line with other studies that estimated the magnitude of the effects for genes and the environment separately for boys and girls (Harlaar et al., 2005)

The twin method assumes that MZ twins are more similar in educational achievement than DZ twins because of their larger genetic resemblance and not because MZ twins are treated more alike, by for example teachers, than DZ twins or experience more similar environments. This equal environment assumption is potentially violated if this similarity in treatment and environment relates to a similarity in a phenotype. Research has shown that the equal environment assumption holds for, amongst others, general cognitive ability and educational achievement (Evans & Martin, 2000; Loehlin, 1989). To be able to generalize the outcome of twin studies to the general population, the twin method further assumes that twins are representative of the general population for the phenotype of interest. Twins do differ from singletons in striking ways with regard to birth conditions. Twins are born, on average, 3-4 weeks prematurely and have approximately 1 kg lower birth weights (Martin et al., 2010). These differences dissipate fairly early on, however, and, already in childhood, twins and singletons have very similar scores for body size (Estourgie-van Burk et al., 2010) and, more relevantly, for general cognitive ability (Webbink et al., 2008) and educational achievement (Cohen et al., 2002; de Zeeuw et al., 2012).

Heritability estimates are always limited to the population in which they have been assessed. The relative contribution of genes and the environment to the variation in educational achievement will be different when either the genetic effects are different or the environmental effects differ, for example, due to differences in SES, national curriculum or educational opportunities. Differences across countries may lead to a relatively larger or smaller role of the environment. The observed heritability estimates for educational achievement in the Netherlands are surprisingly comparable to other countries despite major differences in educational systems and societies. The educational system in The Netherlands is more similar to the one implemented in the UK, both countries have a national curriculum, while the educational system in the USA is more decentralized. A national curriculum could restrict the variation in school environments leading to an increase in the relative contribution of genes to the variation in educational achievement. Differences in heritability for educational achievement between countries might also be explained by differences in society. Several studies have found that the heritability of general cognitive ability is larger in children from middle and upper class families while environmental effects have a larger influence in children from lower income families (Scarr-Salapatek, 1971; Turkheimer et al., 2003). Children from low SES families more often live in bad neighborhoods and are less likely to attend good quality schools. The inequality in income, educational opportunity and circumstances under which children grow up is larger in the USA and the UK compared to the Netherlands. If SES moderates the heritability of educational achievement as it does that of general cognitive ability, a lower percentage of children from disadvantaged groups will lead to a higher heritability of educational achievement.

The common environment had a minimal influence on educational achievement. Nonetheless, the fact that there was evidence for the existence of influence of the common environmental effects on some of the educational domains, just as for general cognitive ability (Deary, Johnson & Houlihan, 2009), is of interest as it is in contrast with what is observed in many other developmental domains, such as personality or psychopathology (Plomin, Asbury & Dunn, 2001). However, growing up with the same parents, SES, attending the same school and even being taught by the same teacher did not seem to contribute much to individual differences between children in educational achievement. The absence of a main effect of the common environments still leaves open the possibility that the school environment indirectly contributes to such individual differences. The same teacher, classroom and peers may be experienced rather differently by children transforming these shared environmental factors into unique environmental effects. Furthermore, there could still be an influence of the common environment due to gene-environment (GxE) interaction, when heritability depends on, for example, SES of the parents (Kendler & Eaves, 1986).

Which aspects of the learning environment make children different in their ability to learn at school still needs to be determined. Earlier research has suggested that the relationship between a student and a teacher might be one of these factors as it plays an important role in motivating children to perform well at school (Turner et al., 2002). A study design using the differences within monozygotic twin pairs can be used to identify these unique environmental effects which have an impact on educational achievement (Asbury, Dunn & Plomin, 2006). Knowledge of the factors in the classroom, which are experienced differently by children, that contribute to differences in the educational achievement may provide opportunities to develop interventions in the school environment to realize each child's learning potential.

TABLE S1 Genetic modeling results for arithmetic, reading, reading comprehension, spelling and educational achievement

		ep	-2ll	df	model	χ2	Δdf	p	a^2	c^2	e^2
Arithmetic											
Grade 1											
0 Saturated		50	23690.61	2915	-	-	-	-			
1 ACE	Boys	12	23722.67	2948	0	32.06	33	.514	.35 [.12-.62]	.29 [.04-.49]	.36 [.30-.43]
	Girls								.71 [.52-.76]	.00 [.00-.13]	.29 [.24-.35]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	23723.73	2949	1	1.06	1	.303	.38 [.15-.63]	.27 [.03-.47]	.35 [.29-.42]
	Girls								.65 [.52-.74]	.06 [.00-.17]	.29 [.24-.36]
3 ACE: Boys = Girls		9	23728.15	2951	2	4.42	2	.110	.60 [.46-.71]	.08 [.00-.20]	.32 [.28-.37]
4 AE: Drop C		***8***	***23729.56***	***2952***	***3***	***1.41***	***1***	***.234***	***.68 [.64-.72]***	-	***.32 [.28-.36]***
Grade 2											
0 Saturated		50	20784.12	2616	-	-	-	-			
1 ACE	Boys	12	20821.89	2649	0	37.76	33	.261	.42 [.15-.69]	.24 [.00-.47]	.34 [.28-.42]
	Girls								.66 [.55-.72]	.01 [.00-.09]	.33 [.27-.40]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	20821.91	2650	1	.02	1	.896	.42 [.15-.69]	.24 [.00-.47]	.34 [.28-.42]
	Girls								.66 [.55-.72]	.01 [.00-.09]	.33 [.27-.40]
3 ACE: Boys = Girls		9	20825.79	2652	2	3.89	2	.143	.66 [.50-.71]	.00 [.00-.13]	.34 [.29-.39]
4 AE: Drop C		***8***	***20825.79***	***2653***	***3***	***.00***	***1***	***1.000***	***.66 [.61-.71]***	-	***.34 [.29-.39]***
Grade 3											
0 Saturated		50	20201.25	2584	-	-	-	-			
1 ACE	Boys	12	20246.89	2617	0	45.63	33	.071	.57 [.32-.70]	.07 [.00-.29]	.36 [.30-.44]
	Girls								.74 [.60-.79]	.00 [.00-.12]	.26 [.21-.32]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	20246.89	2618	1	.00	1	.949	.57 [.32-.70]	.07 [.00-.29]	.36 [.30-.44]
	Girls								.74 [.60-.79]	.00 [.00-.12]	.26 [.21-.32]
3 ACE: Boys = Girls		9	20251.48	2620	2	4.59	2	.101	.69 [.56-.73]	.00 [.00-.12]	.31 [.27-.36]
4 AE: Drop C		***8***	***20251.48***	***2621***	***3***	***.00***	***1***	***1.000***	***.69 [.56-.73]***	***.00 [.00-.12]***	***.31 [.27-.36]***

Grade 4											
0 Saturated		50	14586.79	1943	-	-	-	-			
1 ACE	Boys	12	14641.28	1976	0	54.49	33	.011	.49 [.20-.75]	.20 [.00-.46]	.31 [.24-.39]
	Girls								.56 [.26-.73]	.10 [.00-.37]	.33 [.27-.42]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	14641.66	1977	1	.39	1	.534	.49 [.19-.75]	.21 [.00-.46]	.31 [.24-.39]
	Girls								.67 [.30-.73]	.00 [.00-.34]	.33 [.27-.41]
3 ACE: Boys = Girls		9	14643.19	1979	2	1.53	2	.466	.61 [.42-.73]	.07 [.00-.24]	.32 [.27-.38]
4 AE: Drop C		***8***	***14643.87***	***1980***	***3***	***.678***	***1***	***.410***	***.68 [.63-.73]***	-	***.32 [.27-.37]***
Grade 5											
0 Saturated		50	13128.06	1802	-	-	-	-			
1 ACE	Boys	12	13187.57	1835	0	59.51	33	.003	.75 [.59-.80]	.00 [.00-.14]	.25 [.20-.32]
	Girls								.73 [.49-.79]	.00 [.00-.22]	.27 [.21-.34]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	13187.78	1836	1	.21	1	.645	.75 [.59-.81]	.00 [.00-.14]	.25 [.19-.32]
	Girls								.74 [.50-.79]	.00 [.00-.21]	.26 [.21-.34]
3 ACE: Boys = Girls		9	13188.50	1838	2	.71	2	.700	.74 [.60-.78]	.00 [.00-.13]	.26 [.22-.30]
4 AE: Drop C		***8***	***13188.50***	***1839***	***3***	***.00***	***1***	***1.000***	***.74 [.70-.78]***	-	***.26 [.22-.30]***
Grade 6											
0 Saturated		50	5543.04	757	-	-	-	-			
1 ACE	Boys	12	5610.77	790	0	67.73	33	<.001	.44 [.00-.74]	.20 [.00-.59]	.36 [.25-.51]
	Girls								.68 [.29-.81]	.05 [.00-.40]	.26 [.19-.38]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	5610.81	791	1	.03	1	.853	.44 [.00-.74]	.21 [.00-.58]	.36 [.25-.51]
	Girls								.72 [.30-.81]	02 [.00-.39]	.26 [.19-.38]
3 ACE: Boys = Girls		9	5612.76	793	2	1.95	2	.377	.65 [.37-.76]	.04 [.00-.28]	.31 [.24-.40]
4 AE: Drop C		***8***	***5612.88***	***794***	***3***	***.12***	***1***	***.727***	***.69 [.61-.76]***	-	***.31 [.24-.39]***

Reading											
Grade 2											
0 Saturated		25	14506.08	1689	-	-	-	-			
1 ACE	Boys	9	14515.22	1705	0	9.14	16	.908	.84 [.57-.88]	.00 [.00-.26]	.16 [.12-.21]
	Girls								.83 [.57-.89]	.03 [.00-.29]	.14 [.11-.18]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	8	14515.93	1706	0	.72	1	.396	.78 [.56-.88]	.06 [.00-.27]	.16 [.12-.21]
	Girls								.78 [.56-.89]	.08 [.00-.30]	.14 [.11-.18]
3 ACE: Boys = Girls		6	14516.31	1708	1	.38	2	.827	.78 [.64-.87]	.07 [.00-.21]	.15 [.12-.18]
4 AE: Drop C		***5***	***14517.08***	***1709***	***3***	***.76***	***1***	***.383***	***.85 [.82-.88]***	-	***.15 [.12-.18]***
Grade 3											
0 Saturated		25	15150.08	1755	-	-	-	-			
1 ACE	Boys	9	15164.53	1771	0	14.45	16	.565	.70 [.37-.81]	.06 [.00-.36]	.25 [.19-.32]
	Girls								.66 [.37-.82]	.11 [.00-.39]	.22 [.17-.29]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	8	15164.63	1772	1	.10	1	.750	.75 [.38-.81]	.00 [.00-.35]	.25 [.19-.32]
	Girls								.66 [.37-.82]	.12 [.00-.39]	.22 [.17-.29]
3 ACE: Boys = Girls		6	15165.30	1774	2	.67	2	.714	.73 [.56-.80]	.03 [.00-.19]	.24 [.20-.29]
4 AE: Drop C		***5***	***15165.42***	***1775***	***3***	***.12***	***1***	***.734***	***.77 [.72-.80]***	-	***.23 [.20-.28]***
Grade 4											
0 Saturated		25	11305.09	1332	-	-	-	-			
1 ACE	Boys	9	11326.66	1348	0	21.56	16	.158	.47 [.19-.81]	.31 [.00-.56]	.22 [.16-.30]
	Girls								.80 [.63-.85]	.00 [.00-.15]	.20 [.15-.28]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	8	11326.67	1349	1	.01	1	.935	.48 [.20-.78]	.30 [.14-.55]	.22 [.16-.30]
	Girls								.80 [.71-.85]	.00 [.00-.06]	.20 [.15-.28]
3 ACE: Boys = Girls		6	11331.58	1351	2	4.91	2	.086	.79 [.66-.83]	.00 [.00-.11]	.21 [.17-.26]
4 AE: Drop C		***5***	11331.58	1352	3	.00	1	1.000	.79 [.74-.83]	-	.21 [.17-.26]

Grade 5											
0 Saturated		25	9654.21	1157	-	-	-	-			
1 ACE	Boys	9	9671.31	1173	0	17.10	16	.379	.55 [.28-.87]	.29 [.00-.54]	.16 [.12-.23]
	Girls								.75 [.40-.85]	.05 [.00-.38]	.20 [.15-.28]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	8	9671.34	1174	1	.03	1	.861	.55 [.28-.85]	.28 [.00-.54]	.16 [.12-.23]
	Girls								.78 [.51-.85]	.02 [.00-.28]	.20 [.15-.28]
3 ACE: Boys = Girls		6	9680.20	1176	2	8.86	2	.012	.82 [.67-.85]	.00 [.00-.13]	.18 [.15-.23]
4 AE: Drop C		***5***	***9680.20***	***1177***	***3***	***.00***	***1***	***1.000***	***.82 [.77-.85]***	-	***.18 [.15-.23]***
Reading Comprehension											
Grade 3											
0 Saturated		50	21690.85	2686	-	-	-	-			
1 ACE	Boys	12	21731.89	2719	0	41.04	33	.159	.49 [.23-.70]	.15 [.00-.38]	.36 [.29-.44]
	Girls								.66 [.42-.72]	.00 [.00-.22]	.34 [.28-.41]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	21731.97	2720	1	.09	1	.768	.49 [.24-.70]	.15 [.00-.37]	.35 [.29-.43]
	Girls								.64 [.44-.72]	.02 [.00-.20]	.34 [.28-.42]
3 ACE: Boys = Girls		9	21733.23	2722	2	1.26	2	.533	.63 [.47-.70]	.02 [.00-.16]	.35 [.30-.40]
4 AE: Drop C		***8***	***21733.34***	***2723***	***3***	***.11***	***1***	***.744***	***.65 [.60-.70]***	-	***.35 [.30-.40]***
Grade 4											
0 Saturated		50	17111.12	2122	-	-	-	-			
1 ACE	Boys	12	17164.98	2155	0	53.86	33	.012	.49 [.25-.75]	.23 [.00-.44]	.28 [.22-.35]
	Girls								.63 [.31-.70]	.00 [.00-.29]	.37 [.30-.45]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	17165.00	2156	1	.02	1	.899	.49 [.25-.75]	.23 [.00-.44]	.28 [.22-.35]
	Girls								.61 [.40-.70]	.02 [.00-.21]	.37 [.30-.45]
3 ACE: Boys = Girls		9	17171.34	2158	2	6.34	2	.042	.59 [.43-.72]	.08 [.00-.23]	.32 [.28-.38]
4 AE: Drop C		**8**	**17172.47**	**2159**	**3**	**1.13**	**1**	**.288**	**.68 [.63-.73]**	-	**.32 [.27-.37]**

Grade 5											
0 Saturated		50	16834.20	2045	-	-	-	-			
1 ACE	Boys	12	16922.95	2078	0	88.75	33	<.001	.55 [.27-.73]	.11 [.00-.36]	.33 [.27-.42]
	Girls								.65 [.42-.71]	.00 [.00-.19]	.35 [.29-.44]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	16923.51	2079	1	.57	1	.452	.56 [.29-.73]	.11 [.00-.35]	.33 [.27-.41]
	Girls								.62 [.40-.71]	.03 [.00-.21]	.35 [.29-.44]
3 ACE: Boys = Girls		9	16924.54	2081	2	1.03	2	.597	.60 [.42-.71]	.06 [.00-.21]	.34 [.29-.40]
4 AE: Drop C		***8***	***16925.04***	***2082***	***3***	***.50***	***1***	***.480***	***.66 [.61-.71]***	-	***.34 [.29-.39]***
Grade 6											
0 Saturated		50	12024.86	1427	-	-	-	-			
1 ACE	Boys	12	12060.09	1460	0	35.23	33	.363	.48 [.15-.74]	.19 [.00-.47]	.33 [.25-.43]
	Girls								.45 [.09-.71]	.18 [.00-.49]	.37 [.29-.47]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	12061.18	1461	1	1.09	1	.297	.64 [.12-.75]	.04 [.00-.49]	.32 [.25-.44]
	Girls								.42 [.06-.72]	.21 [.00-.52]	.37 [.29-.48]
3 ACE: Boys = Girls		9	12061.51	1463	2	.33	2	.846	.54 [.32-.70]	.12 [.00-.30]	.35 [.29-.42]
4 AE: Drop C		***8***	***12062.79***	***1464***	***3***	***1.27***	***1***	***.259***	***.66 [.59-.71]***	-	***.34 [.29-.41]***
Spelling											
Grade 1											
0 Saturated		50	21082.01	3049	-	-	-	-			
1 ACE	Boys	12	21122.46	3082	0	40.45	33	.174	.47 [.22-.69]	.17 [.00-.38]	.36 [.30-.44]
	Girls								.30 [.04-.62]	.29 [.00-.52]	.41 [.34-.49]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	21124.26	3083	1	1.80	1	.180	.37 [.17-.53]	.26 [.13-.42]	.37 [.30-.45]
	Girls								.26 [.05-.46]	.34 [.16-.52]	.40 [.34-.48]
3 ACE: Boys = Girls		***9***	***21124.50***	***3085***	***2***	***.23***	***2***	***.890***	***.33 [.18-.47]***	***.28 [.16-.40]***	***.39 [.34-.44]***
4 AE: Drop C		8	21143.02	3086	3	18.52	1	<.001	.64 [.60-.68]	-	.36 [.32-.40]

Grade 2											
0 Saturated		50	18716.77	2783	-	-	-	-			
1 ACE	Boys	12	18749.77	2816	0	33.00	33	.467	.54 [.29-.70]	.09 [.00-.31]	.36 [.30-.44]
	Girls								.66 [.38-.72]	.01 [.00-.27]	.34 [.28-.40]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	18749.93	2817	1	.16	1	.685	.54 [.29-.70]	.10 [.00-.44]	.36 [.30-.44]
	Girls								.61 [.37-.71]	.05 [.00-.27]	.34 [.28-.40]
3 ACE: Boys = Girls		9	18750.27	2819	2	.34	2	.845	.59 [.44-.69]	.06 [.00-.19]	.35 [.31-.40]
4 AE: Drop C		***8***	***18750.99***	***2820***	***3***	***.72***	***1***	***.396***	***.66 [.61-.70]***	-	***.34 [.30-.39]***
Grade 3											
0 Saturated		50	18139.31	2720	-	-	-	-			
1 ACE	Boys	12	18160.87	2753	0	21.56	33	.937	.70 [.47-.76]	.00 [.00-.21]	.30 [.24-.36]
	Girls								.70 [.54-.75]	.00 [.00-.15]	.30 [.25-.37]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	18161.55	2754	1	.68	1	.408	.66 [.45-.75]	.04 [.00-.23]	.30 [.24-.37]
	Girls								.68 [.52-.75]	.01 [.00-.16]	.30 [.25-.37]
3 ACE: Boys = Girls		9	18163.11	2756	2	1.56	2	.459	.70 [.62-.74]	.00 [.00-.06]	.30 [.26-.35]
4 AE: Drop C		***8***	***18163.11***	***2757***	***3***	***.00***	***1***	***1.000***	***.70 [.65-.74]***	-	***.30 [.26-.35]***
Grade 4											
0 Saturated		50	14445.63	2124	-	-	-	-			
1 ACE	Boys	12	14478.83	2157	0	33.19	33	.458	.46 [.18-.73]	.22 [.00-.47]	.32 [.26-.40]
	Girls								.61 [.38-.69]	.00 [.00-.20]	.39 [.31-.48]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	14478.91	2158	1	.09	1	.767	.47 [.19-.73]	.32 [.26-.40]	.32 [.26-.40]
	Girls								.61 [.45-.69]	.01 [.00-.13]	39 [.31-.48]
3 ACE: Boys = Girls		9	14482.87	2160	2	3.95	2	.139	.64 [.46-.70]	.01 [.00-.16]	.35 [.30-.41]
4 AE: Drop C		***8***	***14482.87***	***2161***	***3***	***.00***	***1***	***.945***	***.65 [.59-.70]***	-	***.35 [.30-.41]***

Grade 5											
0 Saturated		50	13199.57	2022	-	-	-	-			
1 ACE	Boys	12	13235.51	2055	0	35.94	33	.333	.68 [.41-.74]	.00 [.00-.24]	.32 [.26-.40]
	Girls								.25 [.03-.51]	.44 [.20-.63]	.31 [.25-.38]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	13235.79	2056	1	.28	1	.594	.63 [.49-.73]	.05 [.00-.16]	.32 [.26-.40]
	Girls								.26 [.04-.51]	.43 [.20-.62]	.31 [.25-.38]
3 ACE: Boys = Girls		9	13245.36	2058	1	9.57	2	.008	.58 [.42-.73]	.11 [.00-.25]	.31 [.27-.37]
4 AE: Boys, Drop C	***Boys***	***10***	***13239.12***	***2057***	***3***	***3.33***	***1***	***.068***	***.69 [.61-.75]***	-	***.31 [.25-.39]***
	Girls								***.38 [.19-.61]***	***.32 [.10-.50]***	***.29 [.24-.36]***
5 AE: Girls, Drop C	Boys	9	13246.87	2058	4	7.75	1	.005	.68 [.60-.74]	-	.32 [.26-.40]
	Girls								.71 [.64-.76]	-	.29 [.24-.36]
Grade 6											
0 Saturated		50	9115.90	1383	-	-	-	-			
1 ACE	Boys	12	9159.94	1416	0	44.04	33	.095	.46 [.15-.75]	.23 [.00-.51]	.30 [.24-.40]
	Girls								.57 [.23-.73]	.09 [.00-.39]	.34 [.27-.44]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	9160.14	1417	1	.19	1	.659	.48 [.15-.75]	.22 [.00-.51]	.30 [.24-.40]
	Girls								.65 [.30-.73]	.01 [.00-.33]	.34 [.27-.43]
3 ACE: Boys = Girls		9	9164.10	1419	2	3.96	2	.138	.66 [.44-.73]	.02 [.00-.21]	.32 [.27-.39]
4 AE: Drop C		***8***	***9164.13***	***1420***	***3***	***.03***	***1***	***.856***	***.68 [.62-.73]***	-	***.32 [.27-.38]***
Educational Achievement											
Arithmetic											
0 Saturated		25	83434.16	8978	-	-	-	-	-	-	-
1 ACE	Boys	9	83448.10	8994	0	13.94	16	.603	.66 [.52-.74]	.05 [.00-.18]	.29 [.26-.33]
	Girls								.68 [.56-.77]	.07 [.00-.19]	.25 [.22-.27]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	8	83449.21	8995	1	1.11	1	.293	.65 [.51-.74]	.06 [.00-.18]	.29 [.26-.33]
	Girls								.75 [.54-.78]	.01 [.00-.20]	.25 [.22-.28]
3 ACE: Boys = Girls		6	83453.08	8997	2	3.87	2	.144	.68 [.61-.75]	.05 [.00-.12]	.27 [.25-.29]
4 AE: Drop C		***5***	***83455.10***	***8998***	***3***	***2.02***	***1***	***.155***	***.74 [.72-.75]***	-	***.26 [.25-.28]***

Language											
0 Saturated		25	83046.16	8978	-	-	-	-	-	-	-
1 ACE	Boys	9	83062.47	8994	0	16.31	16	.431	.62 [.50-.72]	.15 [.05-.26]	.23 [.21-.26]
	Girls								.70 [.60-.76]	.09 [.03-.18]	.22 [.20-.24]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	8	83062.67	8995	1	.20	1	.651	.62 [.50-.72]	.15 [.05-.26]	.23 [.21-.26]
	Girls								.70 [.60-.76]	.09 [.03-.18]	.22 [.20-.24]
3 ACE: Boys = Girls		***6***	***83064.06***	***8997***	***2***	***1.39***	***2***	***.499***	***.67 [.61-.74]***	***.10 [.04-.16]***	***.22 [.21-.24]***
4 AE: Drop C		5	83074.17	8998	2	10.11	1	.001	.78 [.76-.79]	-	.22 [.21-.24]
Study Skills											
0 Saturated		25	83903.97	8969	-	-	-	-	-	-	-
1 ACE	Boys	9	83921.75	8985	0	17.78	16	.337	.51 [.37-.66]	.12 [.00-.25]	.37 [.33-.41]
	Girls								.69 [.56-.72]	.00 [.00-.13]	.31 [.28-.34]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	8	83922.12	8986	1	.37	1	.541	.51 [.37-.65]	.12 [.00-.24]	.37 [.33-.41]
	Girls								.66 [.56-.71]	.03 [.00-.12]	.31 [.28-.34]
3 ACE: Boys = Girls		6	83928.85	8988	2	6.73	2	.035	.60 [.52-.68]	.06 [.00-.13]	.33 [.31-.36]
4 AE: Drop C		***5***	***83931.80***	***8989***	***3***	***2.95***	***1***	***.086***	***.67 [.65-.69]***	-	***.33 [.31-.35]***
Science and Social Studies											
0 Saturated		25	73103.83	7934	-		-	-	-	-	-
1 ACE	Boys	9	73023.23	7950	0	9.41	16	.896	.48 [.36-.61]	.27 [.14-.37]	.25 [.23-.29]
	Girls								.66 [.54-.79]	.13 [.00-.24]	.21 [.19-.23]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	8	73023.25	7951	1	.02	1	.900	.48 [.36-.60]	.27 [.15-.37]	.25 [.23-.29]
	Girls								.65 [.57-.72]	.14 [.07-.22]	.21 [.19-.23]
3 ACE: Boys = Girls		***6***	***73030.43***	***7953***	***2***	***7.18***	***2***	***.028***	***.56 [.50-.64]***	***.21 [.14-.27]***	***.23 [.21-.25]***
4 AE: Drop C		5	73065.31	7954	3	34.88	1	<.001	.78 [.76-.79]	-	.22 [.21-.24]

Total Score											
0 Saturated		25	88844.71	12878	-	-	-	-	-	-	-
1 ACE	Boys	9	88870.41	12894	0	25.70	16	.058	.71 [.61-.81]	.11 [.00-.20]	.18 [.17-.20]
	Girls								.77 [.67-.84]	.06 [.00-.15]	.18 [.16-.19]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	8	88870.60	12895	1	.19	1	.663	.71 [.61-.81]	.11 [.01-.20]	.18 [.17-.20]
	Girls								.79 [.69-.83]	.04 [.00-.13]	.18 [.16-.19]
3 ACE: Boys = Girls		***6***	***88871.41***	***12897***	***2***	***.81***	***2***	***.667***	***.74 [.69-.80]***	***.08 [.02-.13]***	***.18 [.17-.19]***
4 AE: Drop C		5	88879.25	12898	3	7.84	1	.005	.82 [.81-.83]	-	.18 [.17-.19]

REFERENCES

Asbury, K., Dunn, J., & Plomin, R. (2006). The use of discordant MZ twins to generate hypotheses regarding non-shared environmental influence on anxiety in middle childhood. *Social Development, 15 (3),* p. 564-570.

Bartels, M., Rietveld, M. J., van Baal, G. C., & Boomsma, D. I. (2002). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Research, 5,* p. 544-553.

Bartels, M., van Beijsterveldt, C.E., Derks, E.M., Stroet, T.M., Polderman, T.J., Hudziak, J.J., & Boomsma, D.I. (2007). Young Netherlands Twin Register (Y-NTR): a longitudinal multiple informant study of problem behavior. *Twin Research and Human Genetics, 10 (1),* p. 3-11.

Boker S.M., Neale, M. C., Maes, H. H. M., Wilde M.J., Spiegel M., Brick T.R., Spies J., Estabrook R., Kenny S., Bates T.C., Mehta P., & Fox J (2012). OpenMx 1.2 User guide.

Boker S.M., Neale, M. C., Maes, H. H. M., Wilde M.J., Spiegel M., Brick T.R., Estabrook R., Bates T.C., Mehta P., von Oertzen T., Gore R.J., Hunter M.D., Hackett D.C., Karch J., & Brandmaier A. (2011). An open source extended structural equation modeling framework. *Psychometrika, 76 (2),* p. 306-317.

Boomsma, D. I., de Geus, E. J., Vink, J. M., Stubbe, J. H., Distel, M. A., Hottenga, J.J., Posthuma, D., van Beijsterveldt, T.C., Hudziak, J.J., Bartels, M., & Willemsen, G. (2006). Netherlands Twin Register: from twins to twin families. *Twin Research and Human Genetics, 9 (6),* p. 849-857.

Boomsma, D. I., Vink, J. M., van Beijsterveldt, T. C., de Geus, E. J., Beem, A. L., Mulder, E.J., Derks, E.M., Riese, H., Willemsen, G.A., Bartels, M., van den, B.M., Kupper, N.H., Polderman, T.J., Posthuma, D., Rietveld, M.J., Stubbe, J.H., Knol, L.I., Stroet, T., & van Baal, G.C. (2002). Netherlands Twin Register: a focus on longitudinal research. *Twin Research, 5 (5),* p. 401-406.

Boomsma, D. I. (2013). Twin, association and current 'omics' studies. *The Journal of Maternal-Fetal and Neonatal Medicine, 26 (S2),* p. 9-12.

Byrne, B., Coventry, W. L., Olson, R. K., Samuelsson, S., Corley, R., Willcutt, E.G., Wadsworth, S., & DeFries, J.C. (2009). Genetic and environmental influences on aspects of literacy and language in early childhood: Continuity and change from preschool to Grade 2. *Journal of Neurolinguistics, 22 (3),* p. 219-236.

Byrne, B., Wadsworth, S., Boehme, K., Talk, A. C., Coventry, W. L., Olson, R.K., Samuelsson, S., & Corley, R. (2013). Multivariate genetic analysis of learning and early reading development. *Scientific Studies of Reading, 17 (3),* p. 224-242.

Calvin, C. M., Deary, I. J., Webbink, D., Smith, P., Fernandes, C., Lee, S.H., Luciano, M., & Visscher, P.M. (2012). Multivariate genetic analyses of cognition and academic achievement from two population samples of 174,000 and 166,000 school children. *Behavior Genetics, 42 (5),* p. 699-710.

Caravolas, M., Lervåg, A., Defior, S., Málková, G., & Hulme, C. (2013). Different patterns, but equivalent predictors, of growth in reading in consistent and inconsistent orthographies. *Psychological science, 24 (8),* p. 1398-1407.

Chow, B. W.-Y., Ho, C. S.-H., Wong, S. W.-L., Waye, M. M., & Bishop, D. V. (2011). Genetic and environmental influences on Chinese language and reading abilities. *PloS one, 6 (2),* p. e16640.
Cito (2002). Eindtoets Basisonderwijs. Arnhem, The Netherlands: Cito.
Cito (2014a). Cito Leerling Volgsysteem. Arnhem, The Netherlands: Cito.
Cito (2014b). Terugblik en resultaten 2014 - Eindtoets Basisonderwijs. Arnhem, The Netherlands: Cito.
Cohen, C. C., van Goozen, S. H., Orlebeke, J. F., Buitelaar, J. K., & Cohen-Kettenis, P. T. (2002). A comparison of educational achievement in a national sample of Dutch female twins and their matched singleton controls. *Twin Research, 5 (4),* p. 273-276.
de Zeeuw, E. L., van Beijsterveldt, C. E., De Geus, E. J., & Boomsma, D. I. (2012). Twin specific risk factors in primary school achievements. *Twin Research and Human Genetics, 15 (1),* p. 107-115.
Deary, I. J., Johnson, W., & Houlihan, L. M. (2009). Genetic foundations of human intelligence. *Human Genetics, 126 (1),* p. 215-232.
Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35 (1),* p. 13-21.
Estourgie-van Burk, G. F., Bartels, M., Boomsma, D. I., & Delemarre-van de Waal, H. (2010). Body size of twins compared with siblings and the general population: from birth to late adolescence. *The Journal of Pediatrics, 156 (4),* p. 586-591.
Evans, D. M. & Martin, N. G. (2000). The validity of twin studies. *GeneScreen, 1 (2),* p. 77-79.
Frey, M. C. & Detterman, D. K. (2004). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability. *Psychological Science, 15 (6),* p. 373-378.
Göbel, S. M., Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H. C. (2014). Language affects symbolic arithmetic in children: The case of number word inversion. *Journal of Experimental Child Psychology, 119,* p. 17-25.
Harlaar, N., Kovas, Y., Dale, P. S., Petrill, S. A., & Plomin, R. (2012). Mathematics is differentially related to reading comprehension and word decoding: Evidence from a genetically sensitive design. *Journal of Educational Psychology, 104 (3),* p. 622.
Harlaar, N., Spinath, F. M., Dale, P. S., & Plomin, R. (2005). Genetic influences on early word recognition abilities and disabilities: a study of 7-year-old twins. *Journal of Child Psychology and Psychiatry, 46 (4),* p. 373-384.
Hart, S. A., Logan, J. A., Soden-Hensler, B., Kershaw, S., Taylor, J., & Schatschneider, C. (2013). Exploring how nature and nurture affect the development of reading: An analysis of the Florida Twin Project on Reading. *Developmental Psychology, 49 (10),* p. 1971-1981.
Haworth, C. M., Wright, M. J., Luciano, M., Martin, N. G., de Geus, E. J., van Beijsterveldt, C.E., Bartels, M., Posthuma, D., Boomsma, D.I., Davis, O.S., Kovas, Y., Corley, R.P., Defries, J.C., Hewitt, J.K., Olson, R.K., Rhea, S.A., Wadsworth, S.J., Iacono, W.G., McGue, M., Thompson, L.A., Hart, S.A., Petrill, S.A., Lubinski, D., & Plomin, R. (2010). The heritability of general

cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry, 15 (11),* p. 1112-1120.

Haworth, C., Dale, P. S., & Plomin, R. (2009). The etiology of science performance: Decreasing heritability and increasing importance of the shared environment from 9 to 12 years of age. *Child Dsevelopment, 80 (3),* p. 662-673.

Haworth, C., Kovas, Y., Dale, P. S., & Plomin, R. (2008). Science in elementary school: Generalist genes and school environments. *Intelligence, 36 (6),* p. 694-701.

Kendler, K. S. & Eaves, L. J. (1986). Models for the joint effect of genotype and environment on liability to psychiatric illness. *The American Journal of Psychiatry, 143 (3),* p. 279-289.

Kovas, Y., Voronin, I., Kaydalov, A., Malykh, S. B., Dale, P. S., & Plomin, R. (2013). Literacy and numeracy are more heritable than intelligence in primary school. *Psychological Science, 24 (10),* p. 2048-2056.

Loehlin, J. C. (1989). Partitioning environmental and genetic contributions to behavioral development. *American Psychologist, 44 (10),* p. 1285-1292.

Martin, J. A., Hamilton, B. E., Sutton, P. D., Ventura, S. J., Mathews, T. J., Kirmeyer, S., & Osterman, M.J. (2010). Births: final data for 2007. *National Vital Statistics Reports, 58 (24),* p. 1-85.

Moilanen, K. L., Shaw, D. S., & Maxwell, K. L. (2010). Developmental cascades: Externalizing, internalizing, and academic competence from middle childhood to early adolescence. *Development and Psychopathology, 22 (3),* p. 635-653.

Molenaar, D., van der Sluis, S., Boomsma, D. I., Haworth, C. M., Hewitt, J. K., Martin, N.G., Plomin, R., Wright, M.J., & Dolan, C.V. (2013). Genotype by environment interactions in cognitive ability: a survey of 14 studies from four countries covering four age groups. *Behavior Genetics, 43 (3),* p. 208-219.

OECD (2010). PISA 2009 Results: What students know and can do - Student performance in reading, mathematics and science (Volume I).

Plomin R., DeFries J.C., McClearn G.E., & McGuffin P.s (2008). Behavioral Genetics. *5th Edition.* New York, NY: Worth Publishers.

Plomin, R., Asbury, K., & Dunn, J. (2001). Why Are Children in the Same Family So Different? Nonshared Environment a Decade Later. *Canadian Journal of Psychiatry, 46* p. 225-233.

Polderman, T.J., Huizink, A.C., Verhulst, F.C., van Beijsterveldt, C.E., Boomsma, D.I., & Bartels, M. (2011). A genetic study on attention problems and academic skills: results of a longitudinal study in twins. *Journal of the Canadian Academy of Child and Adolescence Psychiatry, 20 (1),* p. 22-34.

Posthuma, D.l., Beem, A.L., De Geus, E.J., Van Baal, G.C., von Hjelmborg, J.B., Iachine, I., & Boomsma, D.I. (2003). Theory and practice in quantitative genetics. *Twin Research, 6 (5),* p. 361-376.

Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research, 5 (6),* p. 554-571.

R Core Team (2014). R: A language and environment for statistical computing. Vieanna, Austria: R Foundation for Statistical Computing.

Rietveld, M. J., van, D., V, Bongers, I. L., Stroet, T. M., Slagboom, P. E., & Boomsma, D. I. (2000). Zygosity diagnosis in young twins by parental report. *Twin Research, 3 (3),* p. 134-141.

Scarr-Salapatek, S. (1971). Race, social class, and IQ. *Science, 54,* p. 260-267.

Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94 (2),* p. 143-174.

Turkheimer, E., Haley, A., Waldron, M., d'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science, 14 (6),* p. 623-628.

Turner, J. C., Midgley, C., Meyer, D. K., Gheen, M., Anderman, E. M., Kang, Y., Patrick, H. (2002). The classroom environment and students' reports of avoidance strategies in mathematics: A multimethod study. *Journal of Educational Psychology, 94 (1),* p. 88-106.

van Beijsterveldt, C. E., Groen-Blokhuis, M., Hottenga, J. J., Franic, S., Hudziak, J. J., ., Lamb, D., Huppertz, C., de Zeeuw, E.L., Nivard, M., Schutte, N., Swagerman, S., Glasner, T., van Fulpen, M., Brouwer, C., Stroet, T., Nowotny, D., Ehli, E.A., Davies, G.E., Scheet, P., Orlebeke, J.F., Kan, K.J., Smit, D., Dolan, C.V., Middeldorp, C.M., de Geus, E.J., Bartels, M., & Boomsma, D.I. (2013). The Young Netherlands Twin Register (YNTR): Longitudinal twin and family studies in over 70,000 children. *Twin Research and Human Genetics, 16 (1),* p. 252-267.

Vlug, K. F. (1997). Because every pupil counts: the success of the pupil monitoring system in The Netherlands. *Education and Information Technologies, 2 (4),* p. 287-306.

Webbink, D., Posthuma, D., Boomsma, D. I., de Geus, E. J., & Visscher, P. M. (2008). Do twins have lower cognitive ability than singletons? *Intelligence, 36 (6),* p. 539-547.

5.

META-ANALYSIS OF TWIN STUDIES HIGHLIGHTS THE IMPORTANTCE OF GENETIC VARIATION IN PRIMARY SCHOOL EDUCATIONAL ACHIEVEMENT

Based on Eveline L. de Zeeuw, Eco J. C. de Geus and Dorret I. Boomsma. Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement. *Trends in Neuroscience and Education*. Submitted.

Children differ in their ability to learn what is taught at school. Evidence from twin studies suggests that genetic effects contribute to such differences. The aim of the present study was to meta-analyze the existing literature on twin studies on educational achievement in primary school children. The meta-analysis includes 61 studies from 11 different cohorts and is based on up to 5330 MZ and 7084 DZ twin pairs. Heritability is estimated at 70% for reading, 50% for reading comprehension, 57% for mathematics, 44% for spelling, and 66% for educational achievement). The importance of genetic effects on educational achievement differed between countries. In general, heritability was consistently high in the Netherlands, while for the USA and UK heritability for some educational domains was moderate or even low. It can be concluded that genetic variation is an important contributor to the individual differences in educational achievement, with evidence for interaction with country.

INTRODUCTION

Educational achievement in children can be defined as the extent to which a child has achieved the educational goals corresponding to his or her grade level. Lower educational achievement has an adverse effect on access to higher education and is negatively related to numerous other outcomes later in life, including earnings (Julian & Kominski, 2011), health and wellbeing (Mackenbach et al., 1997). Research into the causes of individual differences has tended to focus on environmental factors, such as parental educational level, socio-economic status (SES) and quality of education. Yet, even children from a similar background, attending the same school and taught by the same teacher, can differ greatly in their performance at school. This introduces genetic effects as an important additional source of variation in educational achievement. Moreover, parts of the child's environment, like parental educational level, can themselves be influenced by genes (Rietveld et al., 2013; Vinkhuyzen et al., 2010). In keeping, general cognitive ability is the most important predictor of educational achievement (Deary et al., 2007), explaining roughly half of the variation (Frey & Detterman, 2004). A major role for genetic effects on general cognitive ability is well recognized (Plomin, 2004). Here we systematically review twin studies on educational achievement of children in primary school, aiming to provide, based on the existing literature, an estimate of the heritability and the influence of the environment by meta-analyzing the twin correlations.

Twin studies are the most often used design to analyze the causes of variation in complex phenotypes such as educational achievement (Boomsma et al., 2002). Monozygotic (MZ) twin pairs are genetically (nearly) identical while dizygotic

(DZ) twin pairs share approximately 50 per cent of their segregating genes (Plomin R. et al., 2008). If the larger genetic resemblance of MZ twin pairs is mirrored in a larger resemblance for a phenotype, i.e. when the correlation between MZ twin pairs is higher than between DZ twin pairs, this observation is consistent with the phenotype being influenced by genetic effects. Genetic effects are the sum of the effects of all genetic variants with an influence on educational achievement. Environmental effects often are distinguished into common environmental and unique environmental effects. Common environmental effects are influences that are shared between twins or siblings who grow up in the same environment, e.g. SES, and enhance their similarity beyond the similarity due to shared genes. There are other effects that also make offspring from the same parents more similar, including the effects of assortative mating, the similarity between spouses, which will in the classical twin design also be detected as common environment (Evans, Gillespie & Martin, 2002). When the correlation between DZ twin pairs is more than half the correlation between MZ twin pairs there is an indication for the influence of the common environment. Unique environmental effects are influences that are not shared between twins, and make children less similar. When the correlation between MZ twin pairs is not equal to unity the unique environment has an influence. The unique environmental effects also include measurement error, for instance when teacher's reports on achievement test results are incorrect, e.g. wrong child, wrong test.

The twin method assumes that MZ twins are more similar in educational achievement than DZ twins because of their larger genetic resemblance and not because MZ twins are treated more alike than DZ twins. The equal environment assumption can be violated if similarity in treatment relates to similarity in a phenotype, however, MZ twins may be exposed to more similar treatment because of their larger genetic resemblance. For instance, if smart children get treated differently than less smart children, the higher genetic resemblance in cognitive ability of MZ twins causes them to experience more similar environments than DZ twins, as a secondary effect of the genetic effects on cognitive ability. In contrast, when there is a similar environment unrelated to the genetic make-up of the twins, e.g. MZ twins are dressed more alike than DZ twins this could lead to a violation of the assumption, if dress similarity relates to similarity in the outcome. Such violations of the equal environments assumption have been tested by empirical approaches in large scale studies (Evans & Martin, 2000; Loehlin, 1989; van den Oord, Boomsma & Verhulst, 2000) which show that the assumption holds for general cognitive ability, educational achievement and childhood behavioral problems.

In order to generalize the outcome of twin studies to the general population, twins should be representative of the general population for the phenotype of interest. With regard to most characteristics, this assumption will be met as twins are born in all strata of society (Hoekstra et al., 2010). Nonetheless, twins differ from singletons with regard to birth conditions. Twins are born, on average, 3-4 weeks prematurely and have ~1 kg lower birth weights (Martin et al., 2010). These differences dissipate fairly early on, however, and, already in childhood, twins and singletons have very similar body composition (Estourgie-van Burk et al., 2010), general cognitive ability (Webbink et al., 2008) and educational achievement (Cohen et al., 2002), especial when birth order within family is taken into account (Boomsma et al., 2008; de Zeeuw et al., 2012).

The heritability of general cognitive ability as measured by psychometric IQ tests has been studied extensively. A large meta-analysis of twin studies from different countries established that heritability increases linearly from childhood to adulthood from .41 in childhood to .66 in young adulthood. Simultaneously, a decrease from .33 to .16 was seen for the influence of the common environment shared by children growing up in the same family (Haworth et al., 2010; Molenaar et al., 2013). Even during the short period of primary school, heritability of general cognitive ability increased from .38 to .49 (Kovas et al., 2013) and there is a substantial genetic correlation across age (Davis et al., 2008). One explanation that has been suggested for this consistent finding of increasing heritability with age is that children, when they grow up, can more and more select their own environment and experiences based on their genotype (Haworth et al., 2010).

General cognitive ability and educational achievement are positively associated with a correlation of approximately 0.50 (Bartels et al., 2003). Multivariate twin methods (Boomsma, 2014) have been used to analyze the etiology of this association. In childhood, a large part of this association is due to correlated genetic effects, i.e. genes that influence general cognitive ability also influence educational achievement (Bartels et al., 2002; Calvin et al., 2012). General cognitive ability is often seen as a predisposition while educational achievement is perceived as the outcome of education, which leads to expectations that heritability of educational achievement would be lower than for general cognitive ability (Kovas et al., 2013). However, in a study in primary school children literacy and numeracy were more heritable than general cognitive ability (Kovas et al., 2013). One hypothesis for this difference is that the homogeneity of education reduces differences in the environment and, as a result, individual differences between children in educational achievement can to a greater extent be explained by genes (Heath et al., 1985).

Twin studies have mainly focused on reading and, more recently, mathematics. Most studies are from English speaking countries, such as the USA, UK and Australia. Studies from other countries with different educational systems are relatively scarce (Bartels et al., 2002; Byrne et al., 2009; Chow et al., 2011). Studies are characterized by differences in age, sample size, cohort and measurement instrument. Therefore, it is difficult to draw clear conclusions regarding the relative contribution of genetic and environmental influences on educational achievement. Here we aim to provide a review of all studies that addressed the heritability of educational achievement in primary school and carry out a meta-analysis of the correlations in mono- and dizygotic twins. This review does not include twin studies of selected samples (low or high performance) or of learning disabilities, such as dyslexia and dyscalculia, as there are excellent recent reviews (e.g. (Grigorenko, 2004; Schulte-Körne, 2001; Wadsworth, Olson & Defries, 2010; Willcutt et al., 2010)).

METHODS

A search of the published literature was conducted in PubMed to find all relevant papers describing twin studies on the heritability of educational achievement in primary school children published before September 2014. Searches were performed to find any paper in English that contained the words genetics, heritability and twin study combined with educational achievement, educational attainment, school achievement, academic achievement, scholastic achievement, school performance and academic performance as well as with reading, mathematics, arithmetic, spelling and science in its title, key words, abstract or main text. Abstracts of these search results were evaluated and relevant full text articles were retrieved from the internet. The reference lists of all these papers were examined to identify additional studies that had not been located in the initial database search and searches on names of authors who previously published twin studies on educational achievement were performed. Criteria for inclusion were determined a priori and assessed. Only original research reports published in peer-reviewed journals were included in the review. Twin studies including a sample of primary school aged children (6-13 years) were selected. Studies were included when they contained information on heritability estimates for a measure of educational achievement in a specific educational domain, for example, reading or mathematics, or a measure of general educational achievement. Studies were selected when they used standardized tests or teacher assessments to measure educational achievement. Studies reporting on estimates from univariate analyses as well as studies containing univariate estimates from multivariate analyses were included. Only twin studies from unselected genetically sensitive samples were included. From each study, when available, the first author, year of publication, country, cohort, age, sample size, measurement instrument, educational domain and heritability estimates were extracted.

A meta-analysis of studies that provided a description of sample size, with the numbers of MZ and DZ twins, and twin correlations was conducted for both educational achievement in specific educational domains and general educational achievement. The meta-analysis was carried out to estimate heritability across multiple datasets when at least two independent studies from different cohorts reported information on twin correlations and sample size. The decision which study to select and include in the analysis when studies reported twin correlations from the same cohort was based on the largest sample size. The meta-analyses did not make a distinction for gender as almost no studies reported twin correlations separately for boys and girls.

A variance decomposition model was fitted to the twin correlations, weighted by sample size, to estimate the influence of genetic and common environmental effects (Bartels et al., 2003; van Beijsterveldt & Van Baal, 2002; Verweij et al., 2010) on educational achievement using the structural equation modelling program Mx (Neale et al., 1999). With Mx it is possible to analyze the twin correlations from multiple studies in a multi-group analysis and obtain a maximum likelihood estimate of heritability across all studies. It was tested whether the heritability estimate could be constrained to be equal across studies. The difference in goodness of fit between the nested models was assessed with hierarchic chi-squared tests. The difference in the χ^2-statistic is evaluated with the difference in the number of estimated parameters between the nested models as degrees of freedom. A p-value smaller than 0.01 was considered significant.

RESULTS

The PubMed search retrieved 61 studies that were published between 1991 and 2014. Table S1 summarizes the characteristics and results of these twin studies from 6 different, mostly English speaking, countries (mainly Northern Europe, UK and US, but also Australia and China). The studies include heritability estimates for a number of specific educational domains, i.e. reading, reading comprehension, mathematics, spelling, language and science, and general educational achievement. Table S1 gives an overview of the results for the heritability estimates as reported by the included studies. Studies providing separate estimates for the heritability in boys and girls did not report any gender differences (de Zeeuw et al., 2015; Harlaar et al., 2005; Haworth, Dale & Plomin, 2008; Kovas et al., 2007b; Petrill & Thompson, 1994; Reynolds et al., 1996). Some studies used teacher assessments or standardized tests taken at school while others were based on results from tests that had been administered by the researchers through the internet, telephone or during a home-visit. Teacher assessments were based on the evaluation by the teacher of the overall proficiency of a student or on criteria that are listed in national guidelines regarding what a student should be able to do or know regarding a certain educational domain. Some studies took into account that the members of a twin pair could be assessed by the same or different teachers and reported separate heritability estimates for these groups (Harlaar, Dale & Plomin, 2005; Walker et al., 2004).

The cohorts that are described in the studies were the Colorado Learning Disabilities Research Center (CLDRC), UK government's Department of Children, Schools and Families (DSCF), Environmental Risk Longitudinal Twin Study (ERLTS), Florida Twin Project (FTP), International Longitudinal Twin Study (ILTS), Netherlands Twin Register (NTR), Primair Onderwijs en Speciaal Onderwijs Cohort (PRIMA), Virginia Twin Study of Adolescent Behavioral Development (VTSABD) and the Western Reserve Twin Project (WRTP). Most of the studies focused on the so called core educational domains, i.e. reading and mathematics. Other educational domains that we came across in the literature search and that are included in the review are reading comprehension, spelling, language and science. Some of the studies used a measure of general educational achievement. The instruments used to measure educational achievement differ across country and cohort.

Estimates of the heritability of reading (.10-.94), reading comprehension (.32-.87), mathematics (.04-.75), spelling (.33-.84), language (.21-.81), science (.32-.64) and general educational achievement (.27-.57) varied considerably across the studies reported in this review. The same is true for the environmental effects on reading (.00-.74), reading comprehension (.00-.50), mathematics (.00-.81), spelling (.00-.46), language (.10-.25), science (.08-.39) and general educational achievement (.08-.67). Reported heritability estimates may vary due to considerable differences in sample sizes, different countries, different age groups and a large variation in measurement instruments. We explore some of these explanations in the meta-analysis.

A meta-analysis was carried out for reading, reading comprehension, mathematics, spelling and general educational achievement. The MZ and DZ correlations of all studies included in the meta-analyses are given in Table 1. The number of included studies in the meta-analysis was 11 for reading with a total of 5330 MZ and 7084 DZ twin pairs. For reading comprehension a total of 6 studies provided data on 3042 MZ and 5218 DZ twin pairs. For mathematics and spelling, there were fewer studies. Three studies on mathematics included a total of 3419 MZ and 6247 DZ twin pairs and the 3 studies for spelling had 1093 MZ and 1692 DZ twin pairs. In primary school aged children we retrieved 2 studies for general educational achievement with large sample sizes, totaling 4341 MZ and 7808 DZ twin pairs. The heritability estimates reported by the studies included in the meta-analyses and the mean estimate of the heritability based on all available studies are displayed in Figure 1.

We next investigated the heterogeneity between studies for heritability estimates by comparing the fit of the meta-analysis models in which all estimates across studies were constrained to be equal to a model in which all estimates were free. The differences in chi-squared statistics for reading ($\Delta\chi^2 = 25.46$, $\Delta df = 20$, $p = .184$) and general educational achievement ($\Delta\chi^2 = 6.68$, $\Delta df = 2$, $p = .035$) were not significant. For the educational domains reading comprehension ($\Delta\chi^2 = 73.76$, $\Delta df = 14$, $p < .001$), mathematics ($\Delta\chi^2 = 15.58$, $\Delta df = 4$, $p = .004$) and spelling ($\Delta\chi^2 = 30.74$, $\Delta df = 8$, $p < .001$) the constrained model fitted worse, pointing to heterogeneity. The contributions of the included studies to the difference in the chi-squared statistics between the models with all estimates freely estimated and the models were the estimates were constrained to be equal across the different studies are displayed in Table 1 and inform on the degree and sources of heterogeneity across the different samples. Astudy from the Netherlands (de Zeeuw et al., 2015) and a study from the UK (Trzesniewski et al., 2006) both contribute the most to the increase in the chi-square statistic for reading. More than 40 per cent of the increase in the chi-square statistic for reading comprehension is caused by a study in twins from the USA (Hart et al., 2013) and a sample from Australia (Byrne et al., 2009) contributes for nearly half the increase in the chi-square statistic for spelling. The included studies contribute approximately the same to the increase in chi-square statistic for mathematics and general educational achievement.

The studies included in the meta-analyses are mainly from cohorts from the USA, UK and the Netherlands (NL), providing the opportunity to explore gene-environment (GxE) interaction across those countries for the educational domains with studies available from those three countries, i.e. reading, reading comprehension and mathematics (Table 2). These countries have different teaching methods, educational systems and societies and the expression of the genotype could depend on differences in the environment (Eaves, 1984). Heritability and the influence of the common environment, respectively, was first estimated separately for each country. The fit of the model did not deteriorate significantly after equating the estimates across countries for reading ($\Delta\chi^2 = 10.55$, $\Delta df = 4$, $p = .032$), but did so for reading comprehension ($\Delta\chi^2 = 49.80$, $\Delta df = 4$, $p < .001$) and mathematics ($\Delta\chi^2 = 15.58$, $\Delta df = 4$, $p = .004$).

DISCUSSION

The current paper presents a review of the heritability of educational achievement of children in primary school estimated from twin studies. Heritability estimates varied considerably across studies as did the influence of the environmental effects. The small sample sizes, different countries, different age groups and the variety of measurement instruments are probably the main reasons for the broad range of estimates observed in this review. For example, the smallest sample size was 32 MZ and 28 DZ twin pairs (Hohnen & Stevenson, 1999) and the largest was 2292 MZ and 4184 DZ twin pairs (Harlaar, Hayiou-Thomas & Plomin, 2005). It is noteworthy that studies estimating the magnitude of the effects for genes and the environment separately for boys and girls did not find any evidence for quantitative nor qualitative gender differences. This means that in primary school the extent to which genes influence educational achievement is similar across boys and girls and the same genes are involved in educational achievement for boys and girls.

A meta-analysis of twin correlations was performed for reading, reading comprehension, mathematics, spelling and general educational achievement. Many of the studies included in the review used data from the same cohorts. Consequently, the meta-analysis of twin correlations for most educational domains was based on only a few studies. It was not possible to equate the estimates across the studies included in the meta-analyses without a significant drop in model fit for reading comprehension, mathematics and spelling.If we nevertheless averaged the heritability across studies, 73% of the variation in reading, 49% in reading comprehension, 57% in mathematics, 44% in spelling and 66% in general educational achievement could be explained by genetic effects. Common environmental effects explained 10% of the variation in reading, 13% in reading comprehension, 10% in mathematics, 23% in spelling and 12% in general educational achievement. The only selection criteria for the meta-analyses was the largest sample size when studies from the same cohort reported on the same educational achievement domain and this must be kept in mind when evaluating the mean heritability estimates. Overall, the results suggest that educational achievement of the different educational domains is moderate to highly heritable and that the common environment has a small influence.

Further analyses indicated that the heritability of educational achievement in reading comprehension and mathematics, but not reading, is moderated by the country, i.e. USA, UK and the Netherlands, in which children attend school. Heritability of reading was equally high across countries, but heritability of reading comprehension was larger in the Netherlands and the USA compared to the UK and heritability of mathematics was low in the USA, moderate in the UK and large in the Netherlands. Furthermore, the influence of the common environment was larger in the USA and UK compared to the Netherlands. It must be noted that the sample sizes included in the studies from the USA are much smaller, making the estimates less reliable. In general, the heritability estimates are consistently high in the Netherlands while this is not true for the USA and UK. The inequality in educational opportunity, income and circumstances under which children grow up is larger in the USA and the UK compared to the Netherlands. It seems that equal opportunities in the relatively homogenous education environment in the Netherlands reduce environmental variation, making differences in educational achievement between children to a greater extent due to genetic differences. Several studies have already found that the heritability of general cognitive ability is larger in children from middle and upper class families while environmental effects have a larger influence in children from lower income families (Scarr-Salapatek, 1971; Turkheimer et al., 2003).

The consequence of the homogeneity in an educational system is that it will highlight the innate individual differences between children as reflected in the high heritability (Harlaar et al., 2012; Kovas et al., 2013). What must be kept in mind is that this heritability does not equal determinism. The variance between children may be heritable, but the mean can be positively influenced by a school environment of good quality. High heritability in a homogeneous school environment means that children with a predisposition for lower educational achievement will have to struggle while children with a genetic advantage can excel at school without ever tapping their full potential. Heritability does support the role of differentiation in teaching. The double challenge for primary school teachers is to make sure that children, who have more difficulty at school, will learn how to read, write and perform calculations, but that those who have it easy are still sufficiently challenged. Classroom teaching might not be the best method to achieve this goal and a more personalized approach to learning will be necessary.

The next question is whether there is a common set of genes that is influencing educational achievement across different educational domains. A number of studies have demonstrated that a large proportion of the genes that are responsible for the achievements of children in different educational domains are the same (Kovas et al., 2007a). For example, with a genetic correlation of .74 there is a substantial part of the genes with an influence that is shared between mathematics and reading. There is also about one third of the genetic variation that is specific to mathematics and reading (Kovas et al., 2005). The genetic correlation between mathematics and reading comprehension (.76) was significantly larger than between mathematics and word recognition (.50), which suggests that the association with mathematics partly differs between these two components of reading (Harlaar et al., 2012). The genetic correlation between reading and reading comprehension is high, but there are also genetic effects for reading comprehension that are independent from those on reading and vice versa (Betjemann et al., 2008). Although science is less heritable than other educational domains it does share a genetic link with, amongst others, language and mathematics (Haworth et al., 2008). In general, the similarity between the performance of a child in different educational domains is due to genetic rather than environmental effects. Most environmental effects are specific to a certain educational domain and are the cause of individual differences between domains (Kovas et al., 2005).

This is in agreement with the generalist genes hypothesis which holds that many genes associated with one educational domain also influence other domains, that genes associated with educational achievement in the normal range also influence learning disabilities and that genes that influence one aspect of a certain educational domain are largely the same as those that influence other aspects (Plomin & Kovas, 2005). The hypothesis is also supported by multiple studies that have established that learning disabilities are the low end of a continuum and are influenced by the same genetic and environmental effects as normal educational achievement (Hensler et al., 2010; Knopik & DeFries, 1999; Oliver et al., 2004). Heritability estimates of learning disabilities seem to be roughly similar to those for learning abilities (Plomin & Kovas, 2005). Whether high ability is the high end of a continuum of normal variation has been studied less, but seems to be supported for reading (Friend et al., 2009) and mathematics (Petrill et al., 2009).

The same genes are also for a large extent responsible for the performance of children at different ages. Continuity is largely due to the same genetic effects with only some new genes coming into play when a child grows older while environmental effects are responsible for change. For example, heritability of reading at age 7, 9 and 10 was rather similar and the stability across age was primarily genetically medicated with some genes specific to a certain age (Harlaar, Dale & Plomin, 2007b). The longitudinal correlation between mathematics at age 7 and age 9 was for 80 per cent genetically mediated (Haworth et al., 2007). The pattern observed for science is somewhat different since heritability decreased from 9 to 12 years while the shared environmental effects became increasingly important. The genetic correlation of .50 suggests that different genes influence science at these ages (Haworth, Dale & Plomin, 2009).

The phenotypic association between general cognitive ability and educational achievement during the primary school years is largely due to shared genes while differences between the two phenotypes are due to environmental differences. For example, the genetic correlation in a small sample of 6 to 12-year-old twins was very high (.92) while the common unique environmental correlation was only .16 (Petrill & Thompson, 1993). In a study from the Netherlands, the genetic correlation between general cognitive ability and educational achievement in 12-year-olds was somewhat lower (.47) and equal to the unique environmental correlation (Bartels et al., 2002). Genes also explained the largest part of the association between general cognitive ability and specific educational domains, i.e. language and mathematics and to a lesser extent for science (Calvin et al., 2012).

Having established that educational achievement is relatively highly heritable in primary school age children, even more so than general cognitive ability at the same age (Kovas et al., 2013), it is somewhat surprising that no specific genetic variants involved in educational achievement in children have been found. Molecular genetic research towards the lower end of the distribution of reading is most extensive and has yielded promising findings. For recent reviews of the molecular genetic findings for dyslexia see (Carrion-Castillo, Franke & Fisher and Kere (Carrion-Castillo, Franke & Fisher, 2013; Kere, 2014)). In contrast, studies using samples of unselected children are rather scarce and have not yet resulted in conclusive evidence for an association with specific genetic variants. A genome-wide association (GWA) study for reading and spelling including a cohort of 5472 children aged 8 and 9 years from the UK and 1177 older children from Australia (12-25 years) did not find any genetic variants (single nucleotide polymorphisms (SNPs)) associated at a genome-wide significance level. The top results indicated the strongest association with genetic variants in the pseudo gene ABCC13 and the gene DAZAP1. Subsequent gene-based analyses pointed to the genes CD2L1, CDC2L2 and RCAN3 (Luciano et al., 2013). Another GWA study selected the 300 lowest and highest scoring children on mathematics from the 10-year-old TEDS cohort and validated the suggestive associations from this sample in an unselected sample of 2356 children. None of the genetic variants reached genome-wide significance, but genetic variants located within the MMP7, GRIK1 and DNAH5 genes were implicated. The largest effect size observed explained 0.58 per cent of the variance in mathematical performance (Docherty et al., 2010).

The explanation for this lack of significant findings with regard to specific genes influencing educational achievement may be that it is a highly complex phenotype that is caused by many common genetic variants with small effects. The non-significant measured genetic variants in the GWA studies probably did capture relevant genetic variation, but sample sizes have not been large enough to detect these small effects (Flint & Munafo, 2013). This has been confirmed by the observation that polygenic scores including information from all genetic variants, also the non-significant ones, and their effect sizes observed in a meta-analysis of educational attainment in adults actually explained part of the variance in educational achievement in a sample of children (de Zeeuw et al., 2014; Ward et al., 2014).

There are several limitations of this review of the literature about educational achievement in primary school children that should be noted. A rather large number of studies included in the review suffer from a lack of power which has an effect on the reliability of the obtained heritability estimates in these studies. Another limitation is the heterogeneity in the age of the samples and in the measures used to assess educational achievement. Teacher assessments are used to assess educational achievement in some studies while others use objective tests. Although the association between teacher assessments and standardized tests is relatively strong they are likely measuring partly different aspects of a child's educational achievement. Furthermore, the number of studies included in the meta-analyses was rather small compared to the number of studies included in this review due to the fact that many studies were based on the same population cohort.

To summarize, the heritability of educational achievement in primary school was moderate to high with a small influence of the common environment, which means that most environmental effects were unique. There is some indication for GxE interaction for educational achievement across country. The overlap between educational achievement in different educational domains is mainly due to shared genes while the environmental effects are specific per educational domain. Continuity of educational achievement across primary school is mostly due to the same genes while environmental effects are responsible for change. The association between general cognitive ability and educational achievement is largely due to a shared genetic component. Even though conclusive evidence for an association between specific genetic variants and educational achievement has not yet been found, educational achievement across the normal range remains a promising target for molecular genetic research.

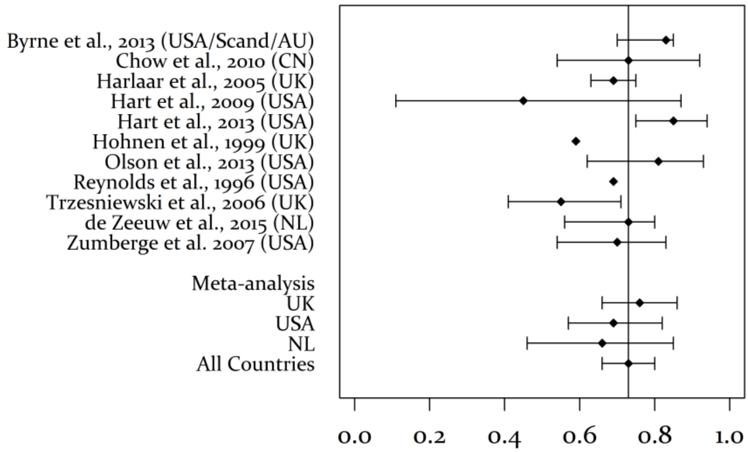
TABLE 1 Descriptives of the studies included in the meta-analyses

Study	Country	Mean Age	Sample Size	MZ correlation	DZ correlation	$\Delta\chi^2$ (df)
Reading						
Byrne et al., 2013	USA Scandinavia Australia	8	433 MZ + 437 DZ	.81	.43	1.51 (2)
Chow et al., 2011	China	-	228 MZ + 84 DZ	.90	.54	1.70 (2)
Harlaar et al., 2005	UK	7	1067 MZ +1039 DZ (boys)	.86	.52	0.59 (6)
			1225 MZ + 1111 DZ (girls)	.84	.51	
			2034 DOS		.37/.46	
Hart et al., 2009	USA	6	128 MZ + 175 DZ	.82	.50	0.45 (2)
Hart et al., 2013	USA	7	486 MZ + 468 DZ	.82	.47	0.23 (3)
			442 DOS		.45	
Hohnen & Stevenson, 1999	UK	7	34 MZ + 32 DZ	.92	.61	1.41 (2)
Olson et al., 2013	USA	11	81 MZ + 189 DZ	.91	.50	0.83 (2)
Reynolds et al., 1996	USA	11	292 MZ + 179 DZ (boys)	.79	.44	1.40 (5)
			380 MZ + 184 DZ (girls)	.81	.52	
			284 DOS		.39	
Trzesniewski et al., 2006	UK	7	285 MZ + 244 DZ	.88	.59	7.81 (2)
de Zeeuw et al., 2015	The Netherlands	8	199 MZ + 182 DZ (boys)	.73	.38	8.49 (5)
			215 MZ + 174 DZ (girls)	.75	.48	
			369 DOS		.39	
Zumberge et al., 2007	USA	10	139 MZ + 84 DZ (boys)	.80	.50	1.05 (5)
			138 MZ + 97 DZ (girls)	.78	.52	
			147 DOS		.43	

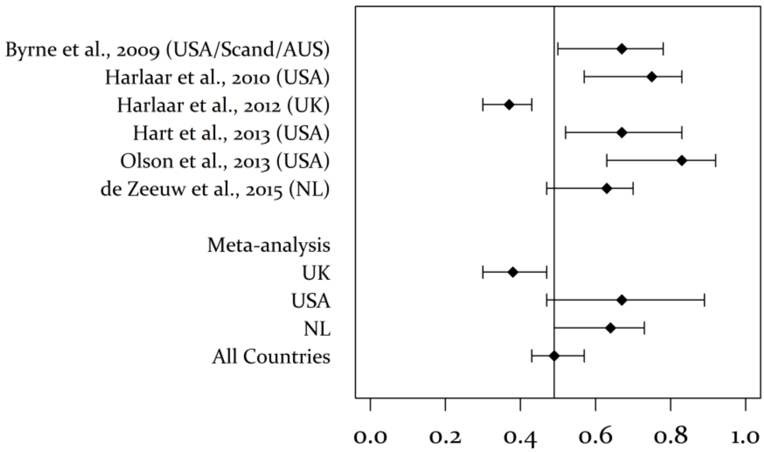
Reading Comprehension						
Byrne et al., 2009	USA	8	185 MZ + 220 DZ	.72	.45	4.88 (2)
	Australia		86 MZ + 49 DZ	.71	.33	1.06 (2)
	Scandinavia		32 MZ + 43 DZ	.76	.46	1.37 (2)
Harlaar et al., 2010	USA	10	89 MZ + 131 DZ	.73	.25	7.37 (2)
Harlaar et al., 2012	UK	12	1748 MZ + 3117 DZ	.56	.37	12.85 (2)
Hart et al., 2013	USA	8	189 MZ + 388 DZ	.83	.53	31.23 (2)
Olson et al., 2013	USA	11	81 MZ + 189 DZ	.88	.47	9.42 (2)
de Zeeuw et al., 2015	The Netherlands	8	305 MZ + 285 DZ (boys)	.67	.41	5.58 (5)
			327 MZ + 261 DZ (girls)	.67	.30	
			535 DOS		.34	
Mathematics						
Harlaar et al., 2012	UK	11	1627 MZ + 2902 DZ	.62	.39	5.52 (2)
Hart et al., 2009	USA	8	128 MZ + 175 DZ	.58	.45	3.47 (2)
de Zeeuw et al., 2015	The Netherlands	12	757 MZ + 787 DZ (boys)	.70	.37	6.59 (5)
			907 MZ + 765 DZ (girls)	.76	.41	
			1618 DOS		.37	
Spelling						
Byrne et al., 2009	USA	8	185 MZ + 220 DZ	.79	.41	3.63 (2)
	Australia		86 MZ + 49 DZ	.74	.10	12.79 (2)
	Scandinavia		32 MZ + 43 DZ	.68	.24	3.92 (2)
Olson et al., 2013	USA	11	81 MZ + 189 DZ	.91	.48	4.96 (2)
de Zeeuw et al., 2015	The Netherlands	6	344 MZ + 316 DZ (boys)	.62	.40	5.43 (5)
			365 MZ + 281 DZ (girls)	.57	.40	
			594 DOS		.53	
Educational Achievement						
Haworth et al., 2011	UK	12	1892 MZ + 3250 DZ	.75	.47	3.85 (2)
de Zeeuw et al., 2015	The Netherlands	12	1112 MZ + 1129 DZ (boys)	.80	.47	2.83 (5)
			1337 MZ + 1149 DZ (girls)	.83	.43	
			2280 DOS		.44	

FIGURE 1 Heritability estimates (95% Confidence Intervals) as reported by the studies included in the meta-analysis and the estimated mean heritability by country for reading (A), reading comprehension (B), mathematics (C), spelling (D) and educational achievement (E)

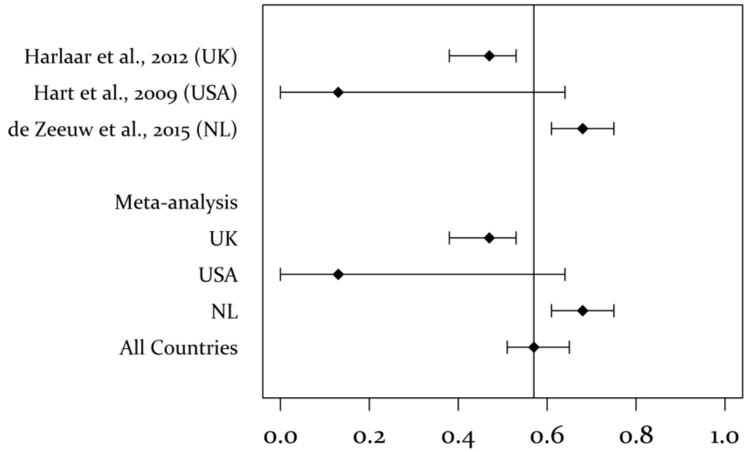
A.



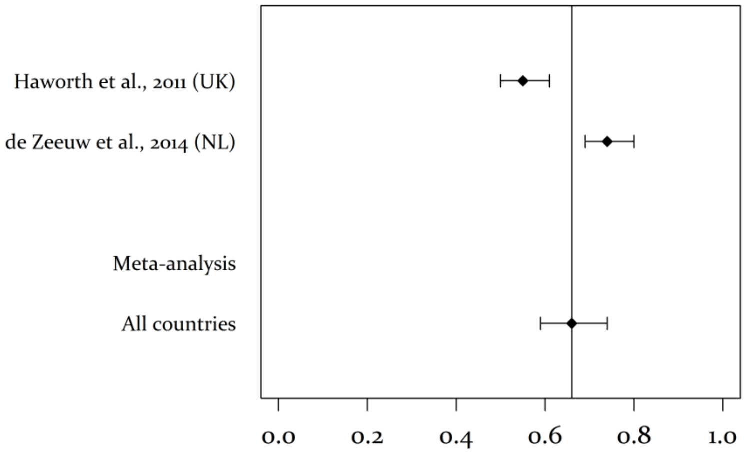
B.



C.


Harlaar et al., 2012 (UK)
Hart et al., 2009 (USA)
de Zeeuw et al., 2015 (NL)
Meta-analysis
UK
USA
NL
All Countries
0.0
0.2
0.4
0.6
0.8
1.0

D.


Haworth et al., 2011 (UK)
de Zeeuw et al., 2014 (NL)
Meta-analysis
All countries
0.0
0.2
0.4
0.6
0.8
1.0

E.

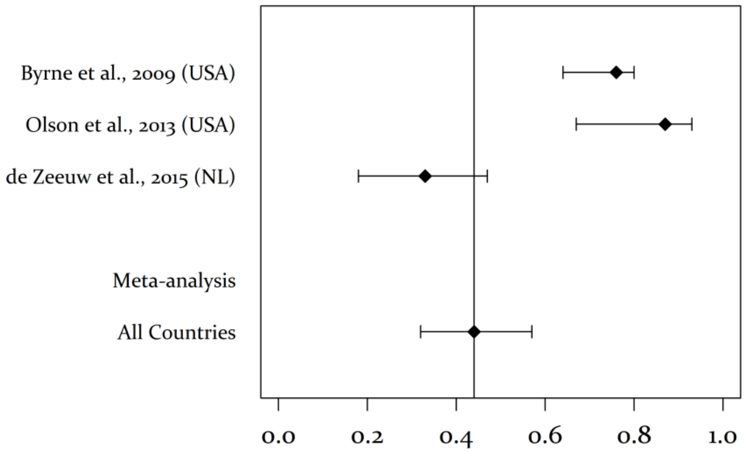

Byrne et al., 2009 (USA)
Olson et al., 2013 (USA)
de Zeeuw et al., 2015 (NL)
Meta-analysis
All Countries
0.0
0.2
0.4
0.6
0.8
1.0

FIGURE 2 Heritability estimates from the meta-analysis of reading, reading comprehension, mathematics, spelling and general educational achievement

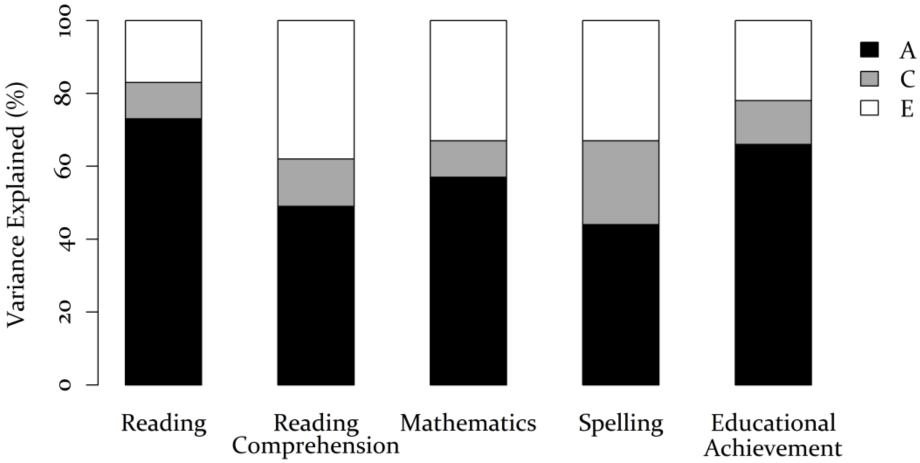


TABLE S1 Twin studies on educational achievement in primary school children

Authors	Country	Study	Study Sample (pairs)	Age (SD) Range	Educational Domain	Test	Subtest	Heritability Estimates
Bartels et al., 2002	The Netherlands	NTR	313 MZ 490 DZ	~12	Educational Achievement	CITO	Total Score	$a^2 = .57$ $c^2 = .27$ $e^2 = .16$
Byrne et al., 2005	USA Australia	ILTS	172 MZ 153 DZ	~6	Reading	TOWRE	-	$a^2 = .70$ $c^2 = .22$ $e^2 = .07$
					Spelling	-	-	$a^2 = .39$ $c^2 = .40$ $e^2 = .20$
Byrne et al., 2006	USA Australia Scandinavia	ILTS	213 MZ 209 DZ		Reading	TOWRE	-	$a^2 = .70$ $c^2 = .22$ $e^2 = .07$
					Spelling	-	-	$a^2 = .39$ $c^2 = .40$ $e^2 = .20$
Byrne et al., 2007	USA Australia	ILTS	167 MZ 152 DZ	~7	Reading	TOWRE	-	$a^2 = .82$ $c^2 = .00$ $e^2 = .18$
					Reading Comprehension	WJ	Passage Comprehension	$a^2 = .76$ $c^2 = .03$ $e^2 = .21$
					Spelling	WJ	Spelling	$a^2 = .71$ $c^2 = .07$ $e^2 = .22$
Byrne et al., 2008	USA Australia	ILTS	225 MZ 214 DZ	8.3 (.4)	Spelling	WRAT	Spelling	$a^2 = .74$ $c^2 = .03$ $e^2 = .23$
Byrne et al., 2009[2,4]	USA Australia Scandinavia	ILTS	303 MZ 312 DZ	~8	Reading	TOWRE	Sight Word	$a^2 = .84$ $c^2 = .00$ $e^2 = .16$
							Phonemic Decoding	$a^2 = .74$ $c^2 = .07$ $e^2 = .19$
							Total	$a^2 = .82$ $c^2 = .03$ $e^2 = .14$
					Reading Comprehension	WJ	-	$a^2 = .67$ $c^2 = .07$ $e^2 = .26$
					Spelling	WRAT	Spelling	$a^2 = .76$ $c^2 = .00$ $e^2 = .24$
Byrne et al., 2013[1]	USA Australia Scandinavia	ILTS	433 MZ 437 DZ	~8	Reading	TOWRE	Sight Word	$a^2 = .83$ $c^2 = .00$ $e^2 = .17$
							Phonemic Decoding	$a^2 = .81$ $c^2 = .01$ $e^2 = .18$

Calvin et al., 2012	UK	DCSF	1056 SS 495 OS	11.2	Educational Achievement	NC-TA	Total	$a^2 = .75$
					Language	NC-TA	English	$a^2 = .81$
					Mathematics	NC-TA	Mathematics	$a^2 = .66$
					Science	NC-TA	Science	$a^2 = .51$
	The Netherlands	PRIMA	785 SS 327 OS	~8	Language	CITO	-	$a^2 = .74$
					Mathematics	CITO	Arithmetic	$a^2 = .67$
				~10	Language	CITO	-	$a^2 = .43$
					Mathematics	CITO	Arithmetic	$a^2 = .73$
				~12	Language	CITO	-	$a^2 = .53$
					Mathematics	CITO	Arithmetic	$a^2 = .36$
Chow et al., 2011[1]	China	-	228 MZ 84 DZ	3-11	Reading	HKT	Word Reading	$a^2 = .73$ $c^2 = .18$ $e^2 = .09$
							Word Attack	$a^2 = .38$ $c^2 = .60$ $e^2 = .02$
Christopher et al., 2013	USA	ILTS	224 MZ 263 DZ	6.3 (.3) 5.5-7.1	Rea ding	TOWRE	Sight Word	$a^2 = .68$ $c^2 = .22$ $e^2 = .10$
							Phonemic Decoding	$a^2 = .62$ $c^2 = .12$ $e^2 = .26$
				7.4 (.3) 6.6-8.7	Reading	TOWRE	Sight Word	$a^2 = .84$ $c^2 = .02$ $e^2 = .14$
							Phonemic Decoding	$a^2 = .82$ $c^2 = .00$ $e^2 = .18$
				8.5 (.3) 7.7-9.5	Reading	TOWRE	Sight Word	$a^2 = .73$ $c^2 = .07$ $e^2 = .20$
							Phonemic Decoding	$a^2 = .40$ $c^2 = .33$ $e^2 = .27$
	Australia	ILTS	152 MZ 115 DZ	6.1 (.4) 5.3-6.8	Reading	TOWRE	Sight Word	$a^2 = .74$ $c^2 = .13$ $e^2 = .13$
							Phonemic Decoding	$a^2 = .83$ $c^2 = .00$ $e^2 = .77$
				7.0 (.4) 6.2-7.8	Reading	TOWRE	Sight Word	$a^2 = .83$ $c^2 = .00$ $e^2 = .17$
							Phonemic Decoding	$a^2 = .60$ $c^2 = .15$ $e^2 = .26$
				8.0 (.4) 7.3-8.7	Reading	TOWRE	Sight Word	$a^2 = .40$ $c^2 = .33$ $e^2 = .27$
							Phonemic Decoding	$a^2 = .81$ $c^2 = .00$ $e^2 = .19$

	Scandinavia	ILTS	138 MZ	6.8 (.3)	Reading	TOWRE	Sight Word	$a^2 = .46$ $c^2 = .43$ $e^2 = .11$
			142 DZ	6.2-7.7			Phonemic Decoding	$a^2 = .42$ $c^2 = .45$ $e^2 = .13$
				7.7 (.3)	Reading	TOWRE	Sight Word	$a^2 = .80$ $c^2 = .03$ $e^2 = .17$
				7.1-8.5			Phonemic Decoding	$a^2 = .67$ $c^2 = .12$ $e^2 = .22$
				8.7 (.3)	Reading	TOWRE	Sight Word	$a^2 = .78$ $c^2 = .00$ $e^2 = .22$
				8.2-9.5			Phonemic Decoding	$a^2 = .71$ $c^2 = .02$ $e^2 = .26$
Christopher et al., 2013	USA	ILTS	210 MZ	6.3	Reading	TOWRE	Sight Word	$a^2 = .68$ $c^2 = .22$ $e^2 = .10$
			254 DZ	5.5-7.1			Phonemic Decoding	$a^2 = .63$ $c^2 = .12$ $e^2 = .25$
				7.4	Reading	TOWRE	Sight Word	$a^2 = .84$ $c^2 = .02$ $e^2 = .14$
				6.6-8.7			Phonemic Decoding	$a^2 = .82$ $c^2 = .00$ $e^2 = .18$
					Reading Comprehension	WRMT	Passage Comprehension	$a^2 = .71$ $c^2 = .10$ $e^2 = .20$
					Spelling	WRAT	Spelling	$a^2 = .68$ $c^2 = .10$ $e^2 = .23$
				8.5	Reading	TOWRE	Sight Word	$a^2 = .73$ $c^2 = .08$ $e^2 = .19$
				7.7-9.5			Phonemic Decoding	$a^2 = .75$ $c^2 = .04$ $e^2 = .21$
					Reading Comprehension	WRMT	Passage Comprehension	$a^2 = .59$ $c^2 = .14$ $e^2 = .27$
					Spelling	WRAT	Spelling	$a^2 = .81$ $c^2 = .00$ $e^2 = .19$
				10.5	Reading	TOWRE	Sight Word	$a^2 = .57$ $c^2 = .15$ $e^2 = .29$
				9.7-11.7			Phonemic Decoding	$a^2 = .74$ $c^2 = .03$ $e^2 = .22$
					Reading Comprehension	WRMT	Passage Comprehension	$a^2 = .68$ $c^2 = .04$ $e^2 = .28$

Davis et al., 2008	UK	TEDS	919 MZ 1622 DZ	~10	Mathematics	NFER	Mathematics	$a^2 = .49$ $c^2 = .23$ $e^2 = .29$
					Reading	PIAT	Reading	$a^2 = .38$ $c^2 = .25$ $e^2 = .37$
Davis et al., 2014	UK	TEDS	2794	~12	Reading	PIAT GOAL TOWRE	Reading Comprehension Reading Comprehension Reading Fluency	$a^2 = .66$ $c^2 = .14$ $e^2 = .20$
					Mathematics	NFER	Numbers Non-Numerical Processes Computation & Knowledge	$a^2 = .51$ $c^2 = .21$ $e^2 = .27$
Gayán and Olson, 2003	USA	CLDRC	257 MZ 183 DZ	10.6 7.8-18.6	Reading	PIAT	Word Recognition	$a^2 = .85$ $c^2 = .04$ $e^2 = .11$
Greven et al., 2009	UK	TEDS	1217 MZ 1070 DZ	9.0 (.3)	Educational Achievement	NC-TA	English Mathematics Science	$a^2 = .70$ $c^2 = .19$ $e^2 = .11$
				10.1 (.3)	Educational Achievement	NC-TA	English Mathematics Science	$a^2 = .51$ $c^2 = .35$ $e^2 = .14$
Greven et al., 2014	UK	TEDS	2191 MZ 3930 DZ	~12	Mathematics	NFER	-	$a^2 = .46$ $c^2 = .23$ $e^2 = .31$
					Reading	TOWRE PIAT WJ GOAL	- Reading Comprehension Reading Fluency -	$a^2 = .56$ $c^2 = .21$ $e^2 = .23$

Hanscombe et al., 2011	UK	TEDS	3843	~12	Language	NC	English	$a^2 = .56$ $c^2 = .25$ $e^2 = .19$
					Mathematics	NC	Mathematics	$a^2 = .49$ $c^2 = .48$ $e^2 = .23$
					Science	NC	Science	$a^2 = .44$ $c^2 = .34$ $e^2 = .22$
Harlaar, Dale, & Plomin, 2005	UK	TEDS	1019 MZ 948 DZ	7.1 (.2)	Reading	TOWRE	Sight Word Phonemic Decoding	same: $a^2 = .63$ $c^2 = .22$ $e^2 = .15$ different: $a^2 = .74$ $c^2 = .09$ $e^2 = .17$
						NC-TA	-	same: $a^2 = .65$ $c^2 = .18$ $e^2 = .17$ different: $a^2 = .66$ $c^2 = .04$ $e^2 = .30$
Harlaar, Hayiou-Thomas, & Plomin, 2005[1]	UK	TEDS	2292 MZ 4184 DZ	7.1 (.2)	Reading	TOWRE	Sight Word Phonemic Decoding	$a^2 = .69$ $c^2 = .16$ $e^2 = .15$
Harlaar et al., 2005	UK	TEDS	1396 MZ 2513 DZ	7.1 (.2)	Reading	TOWRE	-	girls: $a^2 = .67$ $c^2 = .17$ $e^2 = .16$ boys: $a^2 = .65$ $c^2 = .19$ $e^2 = .16$
Harlaar, Dale, & Plomin, 2007	UK	TEDS	1237 MZ 2179 DZ	~7	Reading	NC-TA	-	$a^2 = .67$ $c^2 = .11$ $e^2 = .22$
			899 MZ 1579 DZ	~9	Reading	NC-TA	-	$a^2 = .65$ $c^2 = .10$ $e^2 = .25$
			921 MZ 1651 DZ	~10	Reading	NC-TA	-	$a^2 = .57$ $c^2 = .17$ $e^2 = .26$
Harlaar, Dale, & Plomin, 2007	UK	TEDS	1116 MZ 1923 DZ	~7	Reading	TOWRE	-	$a^2 = .65$ $c^2 = .09$ $e^2 = .15$
				~10	Reading	ART	-	$a^2 = .10$ $c^2 = .49$ $e^2 = .41$
				~12	Reading	TOWRE	-	$a^2 = .67$ $c^2 = .10$ $e^2 = .23$
Harlaar et al., 2010[2]	USA	WRTP	89 MZ 131 DZ	9.9 (.9)	Reading	WRMT	Word Attack	$a^2 = .73$ $c^2 = .11$ $e^2 = .16$
							Word Identification	$a^2 = .81$ $c^2 = .03$ $e^2 = .17$
						TOWRE	Word Decoding	$a^2 = .71$ $c^2 = .10$ $e^2 = .19$
							Sight Word	$a^2 = .78$ $c^2 = .00$ $e^2 = .22$

					Reading Comprehension	WRMT	Passage Comprehension	a^2 = .75 c^2 = .00 e^2 = .25
						PIAT	Reading Comprehension	a^2 = .58 c^2 = .00 e^2 = .42
Harlaar et al., 2012[2]	UK	TEDS	1844 MZ 3318 DZ	11.9 (.6) 11-12	Mathematics	NFER	Understanding Numbers	a^2 = .47 c^2 = .16 e^2 = .37
							Non-Numerical Processes	a^2 = .42 c^2 = .15 e^2 = .44
							Computation & Knowledge	a^2 = .50 c^2 = .08 e^2 = .42
					Reading	TOWRE	Word Reading	a^2 = .74 c^2 = .02 e^2 = .24
						PIAT	Reading Fluency	a^2 = .64 c^2 = .08 e^2 = .28
					Reading Comprehension	PIAT	Reading Comprehension	a^2 = .34 c^2 = .13 e^2 = .53
						GOAL	Reading Comprehension	a^2 = .37 c^2 = .18 e^2 = .45
Hart et al., 2009[1,3]	USA	WRTP	128 MZ 175 DZ	6.1 (.7) 4.3-8.3	Reading	WRMT	Word Identification Passage Comprehension	a^2 = .45 c^2 = .28 e^2 = .19
				7.2 (.7) 6.0-8.8	Reading	WRMT	Word Identification Passage Comprehension	a^2 = .76 c^2 = .07 e^2 = .15
				8.3 (.7) 6.9-10.0	Reading	WRMT	Word Identification Passage Comprehension	a^2 = .81 c^2 = .00 e^2 = .17

						Mathematics	WJ	Calculation	$a^2 = .04$ $c^2 = .50$ $e^2 = .44$
								Fluency	$a^2 = .63$ $c^2 = .15$ $e^2 = .21$
								Applied Problems	$a^2 = .14$ $c^2 = .49$ $e^2 = .37$
								Quantitative Concepts	$a^2 = .00$ $c^2 = .50$ $e^2 = .50$
								WRAT	$a^2 = .13$ $c^2 = .46$ $e^2 = .42$
					9.9 (.8) (8.0-12.1)	Reading	WRMT	Word Identification	$a^2 = .94$ $c^2 = .00$ $e^2 = .12$
								Passage Comprehension	
						Mathematics	WJ	Calculation	$a^2 = .44$ $c^2 = .29$ $e^2 = .26$
								Fluency	$a^2 = .47$ $c^2 = .36$ $e^2 = .14$
								Applied Problems	$a^2 = .54$ $c^2 = .24$ $e^2 = .22$
								Quantitative Concepts	$a^2 = .29$ $c^2 = .52$ $e^2 = .23$
							WRAT	-	$a^2 = .35$ $c^2 = .34$ $e^2 = .32$
Hart, Petrill, & Thompson, 2010	USA		WRTP	94 MZ 134 DZ	9.9 (0.9) 7.4-12.2	Mathematics	WJ	Applied Problems	$a^2 = .41$ $c^2 = .37$ $e^2 = .22$
								Quantitative Concepts	$a^2 = .49$ $c^2 = .32$ $e^2 = .19$
								Calculation	$a^2 = .35$ $c^2 = .39$ $e^2 = .25$
								Fluency	$a^2 = .34$ $c^2 = .46$ $e^2 = .19$
						Reading	WRMT	Word Attack	$a^2 = .74$ $c^2 = .00$ $e^2 = .27$
								Word Identification	$a^2 = .80$ $c^2 = .03$ $e^2 = .17$
							DIBELS	-	$a^2 = .82$ $c^2 = .00$ $e^2 = .17$
							TOWRE	Sight Word Phonemic Decoding	$a^2 = .84$ $c^2 = .00$ $e^2 = .17$
						Reading Comprehension	WRMT	Passage Comprehension	$a^2 = .79$ $c^2 = .00$ $e^2 = .26$

						PIAT	Reading Comprehension	$a^2 = .68$ $c^2 = .00$ $e^2 = .42$
Hart et al., 2010	USA	WRTP	112 MZ 159 DZ	9.8 (1.0)	Reading	WRMT	Word Identification Word Attack	$a^2 = .63$ $c^2 = .35$ $e^2 = .03$
					Mathematics	WJ	Applied Problems Quantitative Concepts Calculation Fluency	$a^2 = .18$ $c^2 = .81$ $e^2 = .01$
Hart et al., 2013a[1]	USA	FTP	824 MZ 1546 DZ	6.7 (.5)	Reading	DIBELS	-	grade 1: $a^2 = .85$ $c^2 = .01$ $e^2 = .14$ grade 2: $a^2 = .63$ $c^2 = .22$ $e^2 = .15$ grade 3: $a^2 = .56$ $c^2 = .28$ $e^2 = .16$ grade 4: $a^2 = .65$ $c^2 = .16$ $e^2 = .19$ grade 5: $a^2 = .71$ $c^2 = .09$ $e^2 = .20$
Hart et al., 2013[2]	USA	FTP	189 MZ 388 DZ	8.2 (1.3)	Reading Comprehension	FCAT	-	$a^2 = .67$ $c^2 = .17$ $e^2 = .15$
Haworth et al., 2007[3]	UK	TEDS	1146 MZ 1032 DZ	7.1 6.5-8.0	Mathematics	NC-TA	Composite	$a^2 = .62$ $c^2 = .16$ $e^2 = .22$
			893 MZ 820 DZ	9.0 8.5-10.5	Mathematics	NC-TA	Using & Applying	$a^2 = .64$ $c^2 = .09$ $e^2 = .27$
							Numbers	$a^2 = .62$ $c^2 = .09$ $e^2 = .29$
							Shapes, Space & Measures	$a^2 = .62$ $c^2 = .11$ $e^2 = .27$
							Composite	$a^2 = .68$ $c^2 = .09$ $e^2 = .23$
Haworth et al., 2008	UK	TEDS	929 MZ 1650 DZ	10.1 (.3)	Language	NC-TA	English	$a^2 = .59$ $c^2 = .20$ $e^2 = .21$
					Mathematics	NC-TA	Mathematics	$a^2 = .62$ $c^2 = .14$ $e^2 = .24$
					Science	NC-TA	Science	$a^2 = .49$ $c^2 = .27$ $e^2 = .24$

Haworth, Dale, & Plomin, 2008	UK	TEDS	1846 MZ 3273 DZ	9.0 8.5-10.5	Science	NC-TA	Composite	$a^2 = .62$ $c^2 = .14$ $e^2 = .23$
Haworth, Dale, & Plomin, 2009	UK	TEDS	910 MZ 1626 DZ	9.0 (.3)	Science	NC-TA	Scientific Enquiry	$a^2 = .56$ $c^2 = .17$ $e^2 = .27$
							Life Processes	$a^2 = .64$ $c^2 = .11$ $e^2 = .25$
							Physical Processes	$a^2 = .62$ $c^2 = .12$ $e^2 = .26$
			1175 MZ 1980 DZ	11.5 (.3)	Science	NC-TA	Scientific Enquiry	$a^2 = .32$ $c^2 = .39$ $e^2 = .29$
							Life Processes	$a^2 = .40$ $c^2 = .33$ $e^2 = .27$
							Physical Processes	$a^2 = .36$ $c^2 = .37$ $e^2 = .27$
							Materials Properties	$a^2 = .40$ $c^2 = .33$ $e^2 = .27$
Haworth et al., 2011	UK	TEDS	1039 MZ 1785 DZ	~12	Educational Achievement	NC-TA	Arithmetic English Science	$a^2 = .52$ $c^2 = .30$ $e^2 = .17$
						NC	Reading Reading Comprehension Mathematics	$a^2 = .55$ $c^2 = .20$ $e^2 = .25$
Hohnen and Stevenson, 1999[1]	UK	-	32 MZ 28 DZ	5.8 5.7-6.1	Reading	BAS SST NARA	- - -	$a^2 = .60$ $c^2 = .36$ $e^2 = .04$
			34 MZ 32 DZ	7.0 6.8-7.3	Reading	BAS SST NARA	- - -	$a^2 = .59$ $c^2 = .32$ $e^2 = .08$

Keenan et al., 2006	USA	CLDRC	70 MZ 61 DZ	~11 8-17	Reading Comprehension	WJ	Passage Comprehension	$a^2 = .51$ $c^2 = .18$ $e^2 = .31$
						QRI	-	
						GORT	-	
						PIAT	Reading Comprehension	
					Reading	PIAT	Word Recognition	$a^2 = .66$ $c^2 = .11$ $e^2 = .23$
Kovas et al., 2005	UK	TEDS	1500 MZ	~7	Mathematics	NC-TA	Mathematics	$a^2 = .67$ $c^2 = .09$ $e^2 = .24$
			1375 DZ		Reading	TOWRE	-	$a^2 = .69$ $c^2 = .17$ $e^2 = .15$
Kovas et al., 2007	UK	TEDS	-	~7	Reading	NC-TA	-	$a^2 = .68$ $c^2 = .07$ $e^2 = .25$
					Reading	TOWRE	Sight Word	$a^2 = .69$ $c^2 = .15$ $e^2 = .17$
							Word Decoding	$a^2 = .67$ $c^2 = .13$ $e^2 = .20$
							Total	$a^2 = .70$ $c^2 = .15$ $e^2 = .15$
					Mathematics	NC-TA	Using & Applying	$a^2 = .65$ $c^2 = .07$ $e^2 = .28$
							Numbers & Algebra	$a^2 = .64$ $c^2 = .06$ $e^2 = .30$
							Shapes, Space & Measures	$a^2 = .66$ $c^2 = .09$ $e^2 = .25$
							Composite	$a^2 = .68$ $c^2 = .09$ $e^2 = .22$
				~9	Reading	NC-TA	-	$a^2 = .64$ $c^2 = .10$ $e^2 = .25$
					Reading Comprehension	PIAT	-	$a^2 = .39$ $c^2 = .25$ $e^2 = .36$
					Mathematics	NC-TA	Using & Applying	$a^2 = .73$ $c^2 = .01$ $e^2 = .26$
							Numbers & Algebra	$a^2 = .67$ $c^2 = .04$ $e^2 = .29$
							Shapes, Space & Measures	$a^2 = .63$ $c^2 = .09$ $e^2 = .28$
							Composite	$a^2 = .72$ $c^2 = .04$ $e^2 = .23$

					Science	NC-TA	Scientific Enquiry	$a^2 = .58$ $c^2 = .13$ $e^2 = .29$
							Life Processes	$a^2 = .65$ $c^2 = .09$ $e^2 = .27$
							Physical Processes	$a^2 = .65$ $c^2 = .08$ $e^2 = .27$
							Composite	$a^2 = .63$ $c^2 = .12$ $e^2 = .24$
				~10	Reading	NC-TA	-	$a^2 = .52$ $c^2 = .20$ $e^2 = .28$
					Mathematics	NC-TA	Using & Applying	$a^2 = .63$ $c^2 = .09$ $e^2 = .28$
							Numbers & Algebra	$a^2 = .62$ $c^2 = .10$ $e^2 = .28$
							Shapes, Space & Measures	$a^2 = .59$ $c^2 = .13$ $e^2 = .28$
							Composite	$a^2 = .64$ $c^2 = .12$ $e^2 = .24$
						NFER	Understanding Numbers	$a^2 = .41$ $c^2 = .18$ $e^2 = .41$
							Non-numerical Processes	$a^2 = .33$ $c^2 = .23$ $e^2 = .44$
							Computation & Knowledge	$a^2 = .46$ $c^2 = .07$ $e^2 = .47$
							Total	$a^2 = .49$ $c^2 = .19$ $e^2 = .32$
					Science	NC-TA	Scientific Enquiry	$a^2 = .48$ $c^2 = .23$ $e^2 = .29$
							Life Processes	$a^2 = .48$ $c^2 = .24$ $e^2 = .28$
							Physical Processes	$a^2 = .45$ $c^2 = .27$ $e^2 = .29$
							Composite	$a^2 = .48$ $c^2 = .27$ $e^2 = .25$
Kovas et al., 2007	UK	TEDS	970 MZ 1704 DZ	~10	Mathematics	NFER	Understanding Numbers	$a^2 = .42$ $c^2 = .16$ $e^2 = .42$
							Non-Numerical Processes	$a^2 = .32$ $c^2 = .23$ $e^2 = .45$
							Computation & Knowledge	$a^2 = .45$ $c^2 = .07$ $e^2 = .48$

Kovas, Petrill, & Plomin, 2007	UK	TEDS	470 MZ 781 DZ	~10	Mathematics	NFER	Mathematical Application	$a^2 = .30$ $c^2 = .30$ $e^2 = .41$
							Understanding Numbers	$a^2 = .42$ $c^2 = .15$ $e^2 = .44$
							Mathematical Interpretation	$a^2 = .42$ $c^2 = .10$ $e^2 = .48$
							Computation & Knowledge	$a^2 = .35$ $c^2 = .08$ $e^2 = .56$
							Non-Numerical Processes	$a^2 = .45$ $c^2 = .11$ $e^2 = .44$
Kovas et al., 2013	UK	TEDS	2415 MZ 2251 DZ	~7	Mathematics	NFER	Numbers	$a^2 = .66$ $c^2 = .16$ $e^2 = .17$
							Using & Applying	
							Shapes, Space & Knowledge	
					Reading	NC-TA	-	$a^2 = .68$ $c^2 = .22$ $e^2 = .10$
						WRMT	Reading Fluency	
						PIAT	Reading Comprehension	
						GOAL	-	
						TOWRE	Total	
			1294 MZ 1152 DZ	~9	Mathematics	NFER	Numbers	$a^2 = .73$ $c^2 = .08$ $e^2 = .19$
							Using & Applying	
							Shapes, Space & Knowledge	
					Reading	NC-TA	-	$a^2 = .77$ $c^2 = .10$ $e^2 = .13$
						WRMT	Reading Fluency	
						PIAT	Reading Comprehension	
						GOAL	-	

			1942 MZ 2192 DZ	~12	Mathematics	NFER	Non-Numerical Processes Understanding Numbers Computation & Knowledge	$a^2 = .56$ $c^2 = .21$ $e^2 = .24$
					Reading	NC-TA WRMT PIAT GOAL TOWRE	- Reading Fluency Passage Comprehension Total -	$a^2 = .65$ $c^2 = .24$ $e^2 = .12$
Logan et al., 2013	USA	WRTP	371 MZ 213 DZ	6.1 (.7) 5.2-7.9 12.2 (1.0) 10.0-14.6	Reading	WRMT	Word Identification	grade 1: $a^2 = .20$ $c^2 = .74$ $e^2 = .06$ grade 2: $a^2 = .33$ $c^2 = .60$ $e^2 = .07$ grade 3: $a^2 = .40$ $c^2 = .50$ $e^2 = .10$ grade 4: $a^2 = .59$ $c^2 = .25$ $e^2 = .16$ grade 5: $a^2 = .66$ $c^2 = .15$ $e^2 = .24$ grade 6: $a^2 = .82$ $c^2 = .00$ $e^2 = .18$
							Word Attack	grade 1: $a^2 = .30$ $c^2 = .58$ $e^2 = .12$ grade 2: $a^2 = .27$ $c^2 = .52$ $e^2 = .21$ grade 3: $a^2 = .58$ $c^2 = .23$ $e^2 = .19$ grade 4: $a^2 = .75$ $c^2 = .00$ $e^2 = .25$ grade 5: $a^2 = .47$ $c^2 = .24$ $e^2 = .29$ grade 6: $a^2 = .71$ $c^2 = .01$ $e^2 = .28$
					Reading Comprehension	WRMT	Reading Comprehension	grade 1: $a^2 = .39$ $c^2 = .39$ $e^2 = .22$ grade 2: $a^2 = .44$ $c^2 = .47$ $e^2 = .09$ grade 3: $a^2 = .32$ $c^2 = .50$ $e^2 = .18$ grade 4: $a^2 = .72$ $c^2 = .05$ $e^2 = .23$ grade 5: $a^2 = .55$ $c^2 = .23$ $e^2 = .12$ grade 6: $a^2 = .82$ $c^2 = .00$ $e^2 = .18$

Luo et al., 2011	UK	TEDS	821 MZ 747 DZ	9.0 (.3)	Mathematics	NC-TA	-	$a^2 = .56$ $c^2 = .12$ $e^2 = .32$
			501 MZ 415 DZ	11.8 (.3)	Mathematics	NC-TA	-	$a^2 = .36$ $c^2 = .17$ $e^2 = .46$
Oliver et al., 2004	UK	TEDS	1146 MZ 1032 DZ	7.1 6.5-8.0	Mathematics	NC-TA	Using & Applying	$a^2 = .61$ $c^2 = .10$ $e^2 = .29$
							Numbers	$a^2 = .63$ $c^2 = .06$ $e^2 = .31$
							Shapes, Space & Measures	$a^2 = .65$ $c^2 = .09$ $e^2 = .26$
							Composite	$a^2 = .66$ $c^2 = .09$ $e^2 = .25$
Oliver, Dale, & Plomin, 2005	UK	TEDS	1565 MZ 1487 DZ	7.2 6.5-8.2	Reading	NC-TA	-	$a^2 = .59$ $c^2 = .18$ $e^2 = .23$
Olson et al., 2011	USA Australia Scandinavia	ILTS	406 MZ 424 DZ	~8	Reading	TOWRE	Word Recognition	$a^2 = .81$ $c^2 = .05$ $e^2 = .14$
							Decoding	$a^2 = .78$ $c^2 = .07$ $e^2 = .25$
					Reading Comprehension	WRMT	Passage Comprehension	$a^2 = .61$ $c^2 = .11$ $e^2 = .27$
	USA	ILTS	176 MZ 213 DZ	10.5 (.3)	Reading	TOWRE	Word Recognition	$a^2 = .77$ $c^2 = .14$ $e^2 = .09$
						WJ	-	
						TOWRE	Phonemic Decoding	$a^2 = .78$ $c^2 = .09$ $e^2 = .14$
					Reading Comprehension	WRMT	Passage Comprehension	$a^2 = .86$ $c^2 = .09$ $e^2 = .04$
Olson et al., 2013[1,2,4]	USA	CLDRC	81 MZ 189 DZ	12.0 (2.9) 8-18	Reading	PIAT	Word Recognition	$a^2 = .81$ $c^2 = .09$ $e^2 = .10$
					Reading Comprehension	WRMT	Reading Comprehension	$a^2 = .87$ $c^2 = .03$ $e^2 = .09$
						PIAT	Reading Comprehension	
					Spelling	WRAT	Spelling	$a^2 = .83$ $c^2 = .05$ $e^2 = .12$
						PIAT	Spelling	

Petrill & Thompson, 1993	USA	WRTP	89 MZ 74 DZ	9.5 (1.8) 6-13	Educational Achievement	MAT	-	$a^2 = .27$ $c^2 = .67$ $e^2 = .06$
Petrill & Thompson, 1994	USA	WRTP	148 MZ 135 DZ	9.6 (1.8) 6-13	Reading	WRAT	-	$a^2 = .55$ $c^2 = .29$ $e^2 = .16$
						MAT	-	$a^2 = .46$ $c^2 = .41$ $e^2 = .13$
					Spelling	WRAT	-	$a^2 = .62$ $c^2 = .22$ $e^2 = .16$
						MAT	-	$a^2 = .53$ $c^2 = .19$ $e^2 = .28$
					Mathematics	WRAT	-	$a^2 = .28$ $c^2 = .51$ $e^2 = .21$
						MAT	-	$a^2 = .36$ $c^2 = .41$ $e^2 = .24$
					Educational Achievement	WRAT	-	$a^2 = .53$ $c^2 = .36$ $e^2 = .11$
						MAT	-	
Petrill et al., 2006	USA	WRTP	102 MZ 140 DZ	6.1 (.7) 4.3-7.9	Reading	WRMT	Word Identification	$a^2 = .68$ $c^2 = .22$ $e^2 = .10$
							Word Attack	$a^2 = .48$ $c^2 = .32$ $e^2 = .20$
Petrill et al., 2006	USA	WRTP	118 MZ 163 DZ	6.1 4-7	Reading	WRMT	Word Identification	$a^2 = .59$ $c^2 = .31$ $e^2 = .10$
Petrill et al., 2007	USA	WRTP	88 MZ 123 DZ	6.1 (.7) 4-7	Reading	WRMT	Word Knowledge	$a^2 = .55$ $c^2 = .34$ $e^2 = .11$
					Reading Comprehension	WRMT	Passage Comprehension	$a^2 = .58$ $c^2 = .33$ $e^2 = .09$
				7.2 (.7) 6-8	Reading	WRMT	Word Knowledge	$a^2 = .50$ $c^2 = .21$ $e^2 = .29$
					Reading Comprehension	WRMT	Passage Comprehension	$a^2 = .76$ $c^2 = .11$ $e^2 = .13$
Reynolds et al., 1996[1]	USA	VTSABD	672 MZ 647 DZ	11.2 (2.5)	Reading	SORT	-	$a^2 = .69$ $c^2 = .13$ $e^2 = .18$
Samuelsson et al., 2007	USA	ILTS	312	~6	Reading	TOWRE	Word Decoding	$a^2 = .49$ $c^2 = .35$ $e^2 = .16$
							Word Recognition	$a^2 = .61$ $c^2 = .33$ $e^2 = .07$
					Spelling	-	-	$a^2 = .34$ $c^2 = .46$ $e^2 = .20$

	Australia	ILTS	157	~6	Reading	TOWRE	Word Decoding	$a^2 = .88$ $c^2 = .00$ $e^2 = .13$
							Word Recognition	$a^2 = .91$ $c^2 = .00$ $e^2 = .09$
					Spelling	-	-	$a^2 = .84$ $c^2 = .00$ $e^2 = .16$
Samuelsson et al., 2008	USA	ILTS	224 MZ 259 DZ	6.3 (.3)	Reading	TOWRE	-	$a^2 = .68$ $c^2 = .25$ $e^2 = .07$
					Spelling	-	-	$a^2 = .44$ $c^2 = .41$ $e^2 = .15$
			185 MZ	7.4 (.3)	Reading	TOWRE	-	$a^2 = .83$ $c^2 = .07$ $e^2 = .11$
			220 DZ		Spelling	-	-	$a^2 = .62$ $c^2 = .13$ $e^2 = .24$
	Australia	ILTS	113 MZ	6.0 (.4)	Reading	TOWRE	-	$a^2 = .84$ $c^2 = .09$ $e^2 = .08$
			78 DZ		Spelling	-	-	$a^2 = .74$ $c^2 = .17$ $e^2 = .10$
			98 MZ	7.0 (.4)	Reading	TOWRE	-	$a^2 = .80$ $c^2 = .02$ $e^2 = .18$
			62 DZ		Spelling	-	-	$a^2 = .79$ $c^2 = .00$ $e^2 = .20$
	Scandinavia	ILTS	65 MZ	6.7 (.3)	Reading	TOWRE	-	$a^2 = .33$ $c^2 = .52$ $e^2 = .15$
			73 DZ		Spelling	-	-	$a^2 = .46$ $c^2 = .42$ $e^2 = .12$
			59 MZ	7.7 (.3)	Reading	TOWRE	-	$a^2 = .79$ $c^2 = .07$ $e^2 = .14$
			66 DZ		Spelling	-	-	$a^2 = .69$ $c^2 = .09$ $e^2 = .22$
Taylor & Schatschneider, 2010	USA	FTP	886 MZ 1684 DZ	6.6 (.4)	Reading	PMRN	Oral Reading Fluency	$a^2 = .62$ $c^2 = .22$ $e^2 = .16$
Thompson, Detterman & Plomin, 1991	USA	WRTP	146 MZ 132 DZ	9.8 6-12	Reading	MAT	-	$a^2 = .29$ $c^2 = .64$ $e^2 = .07$
					Mathematics	MAT	-	$a^2 = .19$ $c^2 = .71$ $e^2 = .10$
					Language	MAT	-	$a^2 = .21$ $c^2 = .62$ $e^2 = .17$
Trzesniewski et al., 2006[1]	UK	ERLTS	285 MZ 244DZ	~7	Reading	TOWRE	-	$a^2 = .55$ $c^2 = .32$ $e^2 = .13$
Walker et al., 2004	UK	TEDS	434 MZ 755 DZ	~7	Educational Achievement	NC-TA	Total	same: $a^2 = .63$ $c^2 = .27$ $e^2 = .10$ different: $a^2 = .42$ $c^2 = .33$ $e^2 = .25$

					Reading	NC-TA	English	same: $a^2 = .63$ $c^2 = .22$ $e^2 = .15$
								different: $a^2 = .37$ $c^2 = .36$ $e^2 = .27$
					Mathematics	NC-TA	Mathematics	same: $a^2 = .69$ $c^2 = .22$ $e^2 = .09$
								different: $a^2 = .46$ $c^2 = .30$ $e^2 = .24$
de Zeeuw et al., 2015[1,2,3,4,5]	The Netherlands	NTR	2451 MZ 4569 DZ	~12	Educational Achievement	CITO	Total Score	$a^2 = .74$ $c^2 = .08$ $e^2 = .18$
					Mathematics	CITO	Arithmetic	$a^2 = .68$ $c^2 = .05$ $e^2 = .27$
					Language	CITO	Language	$a^2 = .67$ $c^2 = .10$ $e^2 = .22$
			1058 MZ 1734 DZ	6-13	Mathematics	CITO	Arithmetic/ Mathematics	grade 1: $a^2 = .60$ $c^2 = .08$ $e^2 = .32$
								grade 2: $a^2 = .66$ $c^2 = .00$ $e^2 = .34$
								grade 3: $a^2 = .69$ $c^2 = .00$ $e^2 = .31$
								grade 4: $a^2 = .61$ $c^2 = .07$ $e^2 = .32$
								grade 5: $a^2 = .74$ $c^2 = .00$ $e^2 = .26$
								grade 6: $a^2 = .65$ $c^2 = .04$ $e^2 = .31$
					Reading	CITO	DMT Reading	grade 2: $a^2 = .78$ $c^2 = .07$ $e^2 = .15$
								grade 3: $a^2 = .73$ $c^2 = .03$ $e^2 = .24$
								grade 4: $a^2 = .79$ $c^2 = .00$ $e^2 = .21$
								grade 5: $a^2 = .82$ $c^2 = .00$ $e^2 = .18$
					Reading Comprehension	CITO	Reading Comprehension	grade 3: $a^2 = .63$ $c^2 = .02$ $e^2 = .35$
								grade 4: $a^2 = .59$ $c^2 = .08$ $e^2 = .32$
								grade 5: $a^2 = .60$ $c^2 = .06$ $e^2 = .34$
								grade 6: $a^2 = .54$ $c^2 = .12$ $e^2 = .35$
					Spelling	CITO	Spelling	grade 1: $a^2 = .33$ $c^2 = .28$ $e^2 = .39$
								grade 2: $a^2 = .59$ $c^2 = .06$ $e^2 = .35$
								grade 3: $a^2 = .70$ $c^2 = .00$ $e^2 = .30$
								grade 4: $a^2 = .64$ $c^2 = .10$ $e^2 = .35$
								grade 5: $a^2 = .58$ $c^2 = .11$ $e^2 = .31$
								grade 6: $a^2 = .66$ $c^2 = .02$ $e^2 = .32$

Zumberge, Baker, & Manis, 2007[2]	USA	-	277 MZ 328 DZ	9.6 (0.6)	Reading	WJ	Word Identification Word Attack	a^2 = .70 c^2 = .11 e^2 = .19

BAS = British Ability Scales; CITO = Dutch National Educational Achievement Test; CLDRC = Colorado Learning Disabilities Research Center; DCSF = UK government's Department of Children, Schools and Families; DIBELS = Dynamic Indicators of Basic Early Literacy Skills; DMT = Three Minute Test; ERLTS = Environmental Risk (E-Risk) Longitudinal Twin Study; FCAT = Florida Comprehensive Achievement Test; FTP = Florida Twin Project; GOAL = Global Online Assessment for Learning; GORT = Gray Oral Reading Test; HKT = Hong Kong Test of Specific Learning Difficulties in Reading and Writing; ILTS = International Longitudinal Twin Study; MAT = Metropolitan Achievement Test; NARA = Neale Analysis of Reading Ability; NC = UK National Curriculum; NC-TA = UK National Curriculum Teacher Assessments; NFER = National Foundation for Educational Research 5–14; NTR = Netherlands Twin Register; PIAT = Peabody Individual Achievement Test; PMRN = Florida's Progress Monitoring and Reporting Network; PRIMA = Primair Onderwijs en Speciaal Onderwijs Cohort; QRI = Qualitative Reading Inventory; TOWRE = Word Reading Efficiency Sight Word Efficiency and Decoding Efficiency; SORT = Slosson Oral Reading Test; SST = Schonell Spelling Test; VTSABD = Virginia Twin Study of Adolescent Behavioral Development; WJ = Woodcock-Johnson III; WRMT = Woodcock Reading Mastery Tests; WRTP = Western Reserve Twin Project; WRAT = Wide-Range Achievement Test

[1] = included in meta-analysis of reading
[2] = included in meta-analysis of reading comprehension
[3] = included in meta-analysis of mathematics
[4] = included in meta-analysis of spelling
[1] = included in meta-analysis of educational achievement

REFERENCES

Bartels, M., Rietveld, M.J., Van Baal, G.C., & Boomsma, D.I. (2002). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Research, 5 (6)*, p. 544-553.

Bartels, M., Van den Berg, M., Sluyter, F., Boomsma, D.I., & de Geus, E.J. (2003). Heritability of cortisol levels: review and simultaneous analysis of twin studies. *Psychoneuroendocrinology, 28 (2)*, p. 121-137.

Betjemann, R.S., Willcutt, E.G., Olson, R.K., Keenan, J.M., DeFries, J.C., & Wadsworth, S.J. (2008). Word reading and reading comprehension: stability, overlap and independence. *Reading and Writing, 21 (5)*, p. 539-558.

Boomsma, D.I., Vink, J.M., van Beijsterveldt, T.C., de Geus, E.J., Beem, A.L., Mulder, E.J., Derks, E.M., Riese, H., Willemsen, G.A., Bartels, M., van den, B.M., Kupper, N.H., Polderman, T.J., Posthuma, D., Rietveld, M.J., Stubbe, J.H., Knol, L.I., Stroet, T., & Van Baal, G.C. (2002). Netherlands Twin Register: a focus on longitudinal research. *Twin Research, 5 (5)*, p. 401-406.

Boomsma, D.I. (2014). Commentary - The strength of multivariate twin studies: testing for shared and distinctive aetiology among different sets of behavioural traits. *Journal of Child Psychology and Psychiatry, 55 (8)*, p. 876-877.

Boomsma, D.I., van Beijsterveld, T.C., Beem, A.L., Hoekstra, R.A., Polderman, T.J., & Bartels, M. (2008). Intelligence and birth order in boys and girls. *Intelligence, 36 (6)*, p. 630-634.

Byrne, B., Coventry, W.L., Olson, R.K., Hulslander, J., Wadsworth, S., DeFries, J.C., Corley, R., Willcutt, E.G., & Samuelsson, S. (2008). A behaviour genetic analysis of orthographic learning, spelling and decoding. *Journal of Research in Reading, 31 (1)*, p. 8-21.

Byrne, B., Coventry, W.L., Olson, R.K., Samuelsson, S., Corley, R., Willcutt, E.G., Wadsworth, S., & DeFries, J.C. (2009). Genetic and environmental influences on aspects of literacy and language in early childhood: continuity and change from preschool to Grade 2. *Journal of Neurolinguistics, 22 (3)*, p. 219-236.

Byrne, B., Olson, R.K., Samuelsson, S., Wadsworth, S., Corley, R., DeFries, J.C., & Willcutt, E. (2006). Genetic and environmental influences on early literacy. *Journal of Research in Reading, 29 (1)*, p. 33-49.

Byrne, B., Samuelsson, S., Wadsworth, S., Hulslander, J., Corley, R., DeFries, J.C., Quain, P., Willcutt, E.G., & Olson, R.K. (2007). Longitudinal twin study of early literacy development: preschool through Grade 1. *Reading and Writing, 20 (1-2)*, p. 77-102.

Byrne, B., Wadsworth, S., Boehme, K., Talk, A.C., Coventry, W.L., Olson, R.K., Samuelsson, S., & Corley, R. (2013). Multivariate genetic analysis of learning and early reading development. *Scientific Studies of Reading, 17 (3)*, p. 224-242.

Byrne, B., Wadsworth, S., Corley, R., Samuelsson, S., Quain, P., DeFries, J.C., Willcutt, E., & Olson, R.K. (2005). Longitudinal twin study of early literacy development: preschool and kindergarten phases. *Scientific Studies of Reading, 9 (3)*, p. 219-235.

Calvin, C.M., Deary, I.J., Webbink, D., Smith, P., Fernandes, C., Lee, S.H., Luciano, M., & Visscher, P.M. (2012). Multivariate genetic analyses of cognition and academic achievement from two population samples of 174,000 and 166,000 school children. *Behavior Genetics, 42 (5)*, p. 699-710.

Carrion-Castillo, A., Franke, B., & Fisher, S.E. (2013). Molecular genetics of dyslexia: an overview. *Dyslexia, 19 (4)*, p. 214-240.

Chow, B.W.-Y., Ho, C.S.-H., Wong, S.W.-L., Waye, M.M., & Bishop, D.V. (2011). Genetic and environmental influences on Chinese language and reading abilities. *PloS One, 6 (2)*, p. e16640

Christopher, M.E., Hulslander, J., Byrne, B., Samuelsson, S., Keenan, J.M., Pennington, B., DeFries, J.C., Wadsworth, S.J., Willcutt, E., & Olson, R.K. (2013a). The genetic and environmental etiologies of individual differences in early reading growth in Australia, the United States, and Scandinavia. *Journal of Experimental Child Psychology, 115 (3)*, p. 453-467.

Christopher, M.E., Hulslander, J., Byrne, B., Samuelsson, S., Keenan, J.M., Pennington, B., DeFries, J.C., Wadsworth, S.J., Willcutt, E.G., & Olson, R.K. (2013b). Modeling the etiology of individual differences in early reading development: evidence for strong genetic influences. *Scientific Studies of Reading, 17 (5)*, p. 350-368.

Cohen, C.C., van Goozen, S.H., Orlebeke, J.F., Buitelaar, J.K., & Cohen-Kettenis, P.T. (2002). A comparison of educational achievement in a national sample of Dutch female twins and their matched singleton controls. *Twin Research, 5 (04)*, p. 273-276.

Davis, O.S.P., Kovas, Y., Harlaar, N., Busfield, P., McMillan, A., Frances, J., Petrill, S.A., Dale, P.S., & Plomin, R. (2008). Generalist genes and the Internet generation: etiology of learning abilities by web testing at age 10. *Genes, Brain and Behavior, 7 (4)*, p. 455-462.

Davis, O.S., Band, G., Pirinen, M., Haworth, C.M., Meaburn, E.L., Kovas, Y., Harlaar, N., Docherty, S.J., Hanscombe, K.B., & Trzaskowski, M. (2014). The correlation between reading and mathematics ability at age twelve has a substantial genetic component. *Nature Communications*, p. 5.

Deary, I.J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35 (1)*, p. 13-21.

Docherty, S.J., Davis, O.S.P., Kovas, Y., Meaburn, E.L., Dale, P.S., Petrill, S.A., Schalkwyk, L.C., & Plomin, R. (2010). A genome-wide association study identifies multiple loci associated with mathematics ability and disability. *Genes, Brain and Behavior, 9 (2)*, p. 234-247.

Estourgie-van Burk, G.F., Bartels, M., Boomsma, D.I., & Delemarre-van de Waal, H. (2010). Body size of twins compared with siblings and the general population: from birth to late adolescence. *The Journal of Pediatrics, 156 (4)*, p. 586-591.

Evans, D.M., Gillespie, N.A., & Martin, N.G. (2002). Biometrical genetics. *Biological Psychology, 61 (1)*, p. 33-51.

Evans, D.M. & Martin, N.G. (2000). The validity of twin studies. *GeneScreen, 1 (2)*, p. 77-79.

Flint, J. & Munafo, M. (2013). Genetics. Herit-ability. *Science, 340 (6139)*, p. 1416-1417.

Frey, M.C. & Detterman, D.K. (2004). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability. *Psychological Science, 15 (6)*, p. 373-378.

Friend, A., DeFries, J.C., Olson, R.K., Pennington, B., Harlaar, N., Byrne, B., Samuelsson, S., Willcutt, E.G., Wadsworth, S.J., & Corley, R. (2009). Heritability of high reading ability and its interaction with parental education. *Behavior Genetics, 39 (4)*, p. 427-436.

Gayán, J. & Olson, R.K. (2003). Genetic and environmental influences on individual differences in printed word recognition. *Journal of Experimental Child Psychology, 84 (2)*, p. 97-123.

Greven, C.U., Kovas, Y., Willcutt, E.G., Petrill, S.A., & Plomin, R. (2014). Evidence for shared genetic risk between ADHD symptoms and reduced mathematics ability: a twin study. *Journal of Child Psychology and Psychiatry, 55 (1)*, p. 39-48.

Greven, C.U., Harlaar, N., Kovas, Y., Chamorro-Premuzic, T., & Plomin, R. (2009). More than just IQ school achievement is predicted by self-perceived abilities - but for genetic rather than environmental reasons. *Psychological Science, 20 (6)*, p. 753-762.

Grigorenko, E.L. (2004). Genetic bases of developmental dyslexia: a capsule review of heritability estimates. *Enfance, 56 (3)*, p. 273-288.

Hanscombe, K.B., Haworth, C., Davis, O.S., Jaffee, S.R., & Plomin, R. (2011). Chaotic homes and school achievement: a twin study. *Journal of Child Psychology and Psychiatry, 52 (11)*, p. 1212-1220.

Harlaar, N., Cutting, L., Deater-Deckard, K., DeThorne, L.S., Justice, L.M., Schatschneider, C., Thompson, L.A., & Petrill, S.A. (2010). Predicting individual differences in reading comprehension: a twin study. *Annals of Dyslexia, 60 (2)*, p. 265-288.

Harlaar, N., Dale, P.S., & Plomin, R. (2005). Telephone testing and teacher assessment of reading skills in 7-year-olds: II. Strong genetic overlap. *Reading and Writing, 18 (5)*, p. 401-423.

Harlaar, N., Dale, P.S., & Plomin, R. (2007a). From learning to read to reading to learn: substantial and stable genetic influence. *Child Development, 78 (1)*, p. 116-131.

Harlaar, N., Dale, P.S., & Plomin, R. (2007b). Reading exposure: a (largely) environmental risk factor with environmentally-mediated effects on reading performance in the primary school years. *Journal of Child Psychology and Psychiatry, 48 (12)*, p. 1192-1199.

Harlaar, N., Hayiou-Thomas, M.E., & Plomin, R. (2005). Reading and general cognitive ability: a multivariate analysis of 7-year-old twins. *Scientific Studies of Reading, 9 (3)*, p. 197-218.

Harlaar, N., Kovas, Y., Dale, P.S., Petrill, S.A., & Plomin, R. (2012). Mathematics is differentially related to reading comprehension and word decoding: evidence from a genetically sensitive design. *Journal of Educational Psychology, 104 (3)*, p. 622-635.

Harlaar, N., Spinath, F.M., Dale, P.S., & Plomin, R. (2005). Genetic influences on early word recognition abilities and disabilities: a study of 7-year-old twins. *Journal of Child Psychology and Psychiatry, 46 (4)*, p. 373-384.

Hart, S.A., Logan, J.A., Soden-Hensler, B., Kershaw, S., Taylor, J., & Schatschneider, C. (2013a). Exploring how nature and nurture affect the development of reading: an analysis of the Florida Twin Project on Reading. *Developmental Psychology, 49 (10)*, p. 1971-1981.

Hart, S.A., Petrill, S.A., & Thompson, L.A. (2010). A factorial analysis of timed and untimed measures of mathematics and reading abilities in school aged twins. *Learning and Individual Differences, 20 (2)*, p. 63-69.

Hart, S.A., Petrill, S.A., Thompson, L.A., & Plomin, R. (2009). The ABCs of math: A genetic analysis of mathematics and its links with reading ability and general cognitive ability. *Journal of Educational Psychology, 101 (2)*, p. 388-402.

Hart, S.A., Petrill, S.A., Willcutt, E., Thompson, L.A., Schatschneider, C., Deater-Deckard, K., & Cutting, L.E. (2010). Exploring how symptoms of attention-deficit/hyperactivity disorder are related to reading and mathematics performance general genes, general environments. *Psychological Science, 21 (11)*, p. 1708-1715.

Hart, S.A., Soden, B., Johnson, W., Schatschneider, C., & Taylor, J. (2013b). Expanding the environment: gene x school-level SES interaction on reading comprehension. *Journal of Child Psychology and Psychiatry, 54 (10)*, p. 1047-1055.

Haworth, C.M., Wright, M.J., Luciano, M., Martin, N.G., de Geus, E.J., van Beijsterveldt, C.E., Bartels, M., Posthuma, D., Boomsma, D.I., Davis, O.S., Kovas, Y., Corley, R.P., Defries, J.C., Hewitt, J.K., Olson, R.K., Rhea, S.A., Wadsworth, S.J., Iacono, W.G., McGue, M., Thompson, L.A., Hart, S.A., Petrill, S.A., Lubinski, D., & Plomin, R. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry, 15 (11)*, p. 1112-1120.

Haworth, C., Dale, P.S., & Plomin, R. (2009). The etiology of science performance: decreasing heritability and increasing importance of the shared environment from 9 to 12 years of age. *Child Development, 80 (3)*, p. 662-673.

Haworth, C., Kovas, Y., Dale, P.S., & Plomin, R. (2008). Science in elementary school: generalist genes and school environments. *Intelligence, 36 (6)*, p. 694-701.

Haworth, C., Kovas, Y., Petrill, S.A., & Plomin, R. (2007). Developmental origins of low mathematics performance and normal variation in twins from 7 to 9 years. *Twin Research and Human Genetics, 10 (1)*, p. 106-117.

Haworth, C.M., Asbury, K., Dale, P.S., & Plomin, R. (2011). Added value measures in education show genetic as well as environmental influence. *PloS one, 6 (2)*, p. e16006.

Haworth, C.M., Dale, P., & Plomin, R. (2008). A twin study into the genetic and environmental influences on academic performance in science in nine-year-old boys and girls. *International Journal of Science Education, 30 (8)*, p. 1003-1025.

Heath, A.C., Berg, K., Eaves, L.J., Solaas, M.H., Corey, L.A., Sundet, J., Magnus, P., & Nance, W.E. (1985). Education policy and the heritability of educational attainment. *Nature, 314* p. 734-736.

Hensler, B.S., Schatschneider, C., Taylor, J., & Wagner, R.K. (2010). Behavioral genetic approach to the study of dyslexia. *Journal of Developmental and Behavioral Pediatrics, 31 (7)*, p. 525-532.

Hoekstra, C., Willemsen, G., van Beijsterveldt, C.E.M., Lambalk, C.B., Montgomery, G.W., & Boomsma, D.I. (2010). Body composition, smoking, and spontaneous dizygotic twinning. *Fertility and Sterility, 93 (3)*, p. 885-893.

Hohnen, B. & Stevenson, J. (1999). The structure of genetic influences on general cognitive, language, phonological, and reading abilities. *Developmental Psychology, 35 (2)*, p. 590-603.

Julian, T. & Kominski, R. (2011). Education and synthetic work-life earnings estimates. American Community Survey Report: Census.

Keenan, J.M., Betjemann, R.S., Wadsworth, S.J., DeFries, J.C., & Olson, R.K. (2006). Genetic and environmental influences on reading and listening comprehension. *Journal of Research in Reading, 29 (1)*, p. 75-91.

Kere, J. (2014). The molecular genetics and neurobiology of developmental dyslexia as model of a complex phenotype. *Biochemical and Biophysical Research Communications, 452 (2)*, p. 236-243.

Knopik, V.S. & DeFries, J.C. (1999). Etiology of covariation between reading and mathematics performance: a twin study. *Twin Research, 2 (3)*, p. 226-234.

Kovas, Y., Voronin, I., Kaydalov, A., Malykh, S.B., Dale, P.S., & Plomin, R. (2013). Literacy and numeracy are more heritable than intelligence in primary school. *Psychoogical Science, 24 (10)*, p. 2048-2056.

Kovas, Y., Harlaar, N., Petrill, S.A., & Plomin, R. (2005). Generalist genes and mathematics in 7-year-old twins. *Intelligence, 33 (5)*, p. 473-489.

Kovas, Y., Haworth, C.M., Dale, P.S., & Plomin, R. (2007a). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for research in Child Development, 72 (3)*, p. 1-144.

Kovas, Y., Haworth, C.M., Petrill, S.A., & Plomin, R. (2007b). Mathematical ability of 10-year-old boys and girls: genetic and environmental etiology of typical and low performance. *Journal of Learning Disabilities, 40 (6)*, p. 554-567.

Kovas, Y., Petrill, S.A., & Plomin, R. (2007). The origins of diverse domains of mathematics: generalist genes but specialist environments. *Journal of Educational Psychology, 99 (1)*, p. 128-139.

Loehlin, J.C. (1989). Partitioning environmental and genetic contributions to behavioral development. *American Psychologist, 44 (10)*, p. 1285

Logan, J.A., Hart, S.A., Cutting, L., Deater-Deckard, K., Schatschneider, C., & Petrill, S. (2013). Reading development in young children: genetic and environmental influences. *Child Development, 84 (6)*, p. 2131-2144.

Luciano, M., Evans, D.M., Hansell, N.K., Medland, S.E., Montgomery, G.W., Martin, N.G., Wright, M.J., & Bates, T.C. (2013). A genome-wide association study for reading and language abilities in two population cohorts. *Genes, Brain and Behavior, 12 (6)*, p. 645-652.

Luo, Y.L., Kovas, Y., Haworth, C.M., & Plomin, R. (2011). The etiology of mathematical self-evaluation and mathematics achievement: understanding

the relationship using a cross-lagged twin study from age 9 to 12. *Learning and Individual Differences, 21 (6)*, p. 710-718.
Mackenbach, J.P., Kunst, A.E., Cavelaars, A.E., Groenhof, F., & Geurts, J.J. (1997). Socioeconomic inequalities in morbidity and mortality in western Europe. The EU Working Group on Socioeconomic Inequalities in Health. *Lancet, 349 (9066)*, p. 1655-1659.
Martin, J.A., Hamilton, B.E., Sutton, P.D., Ventura, S.J., Mathews, T.J., Kirmeyer, S., & Osterman, M.J. (2010). Births: final data for 2007. *National Vitality Statistics Report, 58 (24)*, p. 1-85.
Molenaar, D., van der Sluis, S., Boomsma, D.I., Haworth, C.M., Hewitt, J.K., Martin, N.G., Plomin, R., Wright, M.J., & Dolan, C.V. (2013). Genotype by environment interactions in cognitive ability: a survey of 14 studies from four countries covering four age groups. *Behavior Genetics, 43 (3)*, p. 208-219.
Neale, M.C., Boker, S.M., Xie G., & Maes.H.H. (1999). Mx: Statistical Modeling. Richmond, VA: Department of Psychiatry.
Oliver, B., Harlaar, N., Hayiou Thomas, M.E., Kovas, Y., Walker, S.O., Petrill, S.A., Spinath, F.M., Dale, P.S., & Plomin, R. (2004). A twin study of teacher-reported mathematics performance and low performance in 7-year-olds. *Journal of Educational Psychology, 96 (3)*, p. 504-517.
Oliver, B.R., Dale, P.S., & Plomin, R. (2005). Predicting literacy at age 7 from preliteracy at age 4: a longitudinal genetic analysis. *Psychological Science, 16 (11)*, p. 861-865.
Olson, R.K., Hulslander, J., Christopher, M., Keenan, J.M., Wadsworth, S.J., Willcutt, E.G., Pennington, B.F., & DeFries, J.C. (2013). Genetic and environmental influences on writing and their relations to language and reading. *Annals of Dyslexia, 63 (1)*, p. 25-43.
Olson, R.K., Keenan, J.M., Byrne, B., Samuelsson, S., Coventry, W.L., Corley, R., Wadsworth, S.J., Willcutt, E.G., DeFries, J.C., & Pennington, B.F. (2011). Genetic and environmental influences on vocabulary and reading development. *Scientific Studies of Reading, 15 (1)*, p. 26-46.
Petrill, S.A., Deater-Deckard, K., Thompson, L.A., De Thorne, L.S., & Schatschneider, C. (2006a). Reading skills in early readers genetic and shared environmental influences. *Journal of Learning Disabilities, 39 (1)*, p. 48-55.
Petrill, S.A., Deater-Deckard, K., Thompson, L.A., DeThorne, L.S., & Schatschneider, C. (2006b). Genetic and environmental effects of serial naming and phonological awareness on early reading outcomes. *Journal of Educational Psychology, 98 (1)*, p. 112-121.
Petrill, S.A., Deater-Deckard, K., Thompson, L.A., Schatschneider, C., DeThorne, L.S., & Vandenbergh, D.J. (2007). Longitudinal genetic analysis of early reading: the Western Reserve reading project. *Reading and Writing, 20 (1-2)*, p. 127-146.
Petrill, S.A., Kovas, Y., Hart, S.A., Thompson, L.A., & Plomin, R. (2009). The genetic and environmental etiology of high math performance in 10-year-old twins. *Behavior Genetics, 39 (4)*, p. 371-379.

Petrill, S.A. & Thompson, L.A. (1993). The phenotypic and genetic relationships among measures of cognitive ability, temperament, and scholastic achievement. *Behavior Genetics, 23 (6)*, p. 511-518.
Petrill, S.A. & Thompson, L.A. (1994). The effect of gender upon heritability and common environmental estimates in measures of scholastic achievement. *Personality and Individual Differences, 16 (4)*, p. 631-640.
Plomin R., DeFries J.C., McClearn G.E., & McGuffin P.s (2008). Behavioral genetics. *5th Edition.* New York, NY: Worth Publishers.
Plomin, R. & Kovas, Y. (2005). Generalist genes and learning disabilities. *Psychological Bulletin, 131 (4)*, p. 592-617.
Plomin, R. (2004). Genetics and developmental psychology. *Merrill-Palmer Quarterly, 50 (3)*, p. 341-352.
Reynolds, C.A., Hewitt, J.K., Erickson, M.T., Silberg, J.L., Rutter, M., Simonoff, E., Meyer, J., & Eaves, L.J. (1996). The genetics of children's oral reading performance. *Journal of Child Psychology and Psychiatry, 37 (4)*, p. 425-434.
Rietveld, C.A., Medland, S.E., Derringer, J., Yang, J., Esko, T., Martin, N.W., Westra, H.J., Shakhbazov, K., Abdellaoui, A., Agrawal, A., Albrecht, E., Alizadeh, B.Z., Amin, N., Barnard, J., Baumeister, S.E., Benke, K.S., Bielak, L.F., Boatman, J.A., Boyle, P.A., Davies, G., de, L.C., Eklund, N., Evans, D.S., Ferhmann, R., Fischer, K., Gieger, C., Gjessing, H.K., Hagg, S., Harris, J.R., Hayward, C., Holzapfel, C., Ibrahim-Verbaas, C.A., Ingelsson, E., Jacobsson, B., Joshi, P.K., Jugessur, A., Kaakinen, M., Kanoni, S., Karjalainen, J., Kolcic, I., Kristiansson, K., Kutalik, Z., Lahti, J., Lee, S.H., Lin, P., Lind, P.A., Liu, Y., Lohman, K., Loitfelder, M., McMahon, G., Vidal, P.M., Meirelles, O., Milani, L., Myhre, R., Nuotio, M.L., Oldmeadow, C.J., Petrovic, K.E., Peyrot, W.J., Polasek, O., Quaye, L., Reinmaa, E., Rice, J.P., Rizzi, T.S., Schmidt, H., Schmidt, R., Smith, A.V., Smith, J.A., Tanaka, T., Terracciano, A., van der Loos, M.J., Vitart, V., Volzke, H., Wellmann, J., Yu, L., Zhao, W., Allik, J., Attia, J.R., Bandinelli, S., Bastardot, F., Beauchamp, J., Bennett, D.A., Berger, K., Bierut, L.J., Boomsma, D.I., Bultmann, U., Campbell, H., Chabris, C.F., Cherkas, L., Chung, M.K., Cucca, F., de, A.M., De Jager, P.L., De Neve, J.E., Deary, I.J., Dedoussis, G.V., Deloukas, P., Dimitriou, M., Eiriksdottir, G., Elderson, M.F., Eriksson, J.G., Evans, D.M., Faul, J.D., Ferrucci, L., Garcia, M.E., Gronberg, H., Guethnason, V., Hall, P., Harris, J.M., Harris, T.B., Hastie, N.D., Heath, A.C., Hernandez, D.G., Hoffmann, W., Hofman, A., Holle, R., Holliday, E.G., Hottenga, J.J., Iacono, W.G., Illig, T., Jarvelin, M.R., Kahonen, M., Kaprio, J., Kirkpatrick, R.M., Kowgier, M., Latvala, A., Launer, L.J., Lawlor, D.A., Lehtimaki, T., Li, J., Lichtenstein, P., Lichtner, P., Liewald, D.C., Madden, P.A., Magnusson, P.K., Makinen, T.E., Masala, M., McGue, M., Metspalu, A., Mielck, A., Miller, M.B., Montgomery, G.W., Mukherjee, S., Nyholt, D.R., Oostra, B.A., Palmer, L.J., Palotie, A., Penninx, B.W., Perola, M., Peyser, P.A., Preisig, M., Raikkonen, K., Raitakari, O.T., Realo, A., Ring, S.M., Ripatti, S., Rivadeneira, F., Rudan, I., Rustichini, A., Salomaa, V., Sarin, A.P., Schlessinger, D., Scott, R.J., Snieder, H., St, P.B., Starr, J.M., Sul, J.H., Surakka, I., Svento, R., Teumer, A., Tiemeier, H., van Rooij, F.J., Van Wagoner, D.R., Vartiainen, E., Viikari, J., Vollenweider, P., Vonk, J.M., Waeber, G., Weir, D.R., Wichmann, H.E., Widen, E., Willemsen, G., Wilson,

J.F., Wright, A.F., Conley, D., Davey-Smith, G., Franke, L., Groenen, P.J., Hofman, A., Johannesson, M., Kardia, S.L., Krueger, R.F., Laibson, D., Martin, N.G., Meyer, M.N., Posthuma, D., Thurik, A.R., Timpson, N.J., Uitterlinden, A.G., van Duijn, C.M., Visscher, P.M., Benjamin, D.J., Cesarini, D., & Koellinger, P.D. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science, 340 (6139)*, p. 1467-1471.

Samuelsson, S., Byrne, B., Olson, R.K., Hulslander, J., Wadsworth, S., Corley, R., Willcutt, E.G., & DeFries, J.C. (2008). Response to early literacy instruction in the United States, Australia, and Scandinavia: a behavioral-genetic analysis. *Learning and Individual Differences, 18 (3)*, p. 289-295.

Samuelsson, S., Olson, R., Wadsworth, S., Corley, R., DeFries, J.C., Willcutt, E., Hulslander, J., & Byrne, B. (2007). Genetic and environmental influences on prereading skills and early reading and spelling development in the United States, Australia, and Scandinavia. *Reading and Writing, 20 (1-2)*, p. 51-75.

Scarr-Salapatek, S. (1971). Race, social class, and IQ. *Science, 174 (4016)*, p. 1285-1295.

Schulte-Körne, G. (2001). Annotation - Genetics of reading and spelling disorder. *Journal of Child Psychology and Psychiatry, 42 (8)*, p. 985-997.

Taylor, J. & Schatschneider, C. (2010). Genetic influence on literacy constructs in kindergarten and first grade: evidence from a diverse twin sample. *Behavior Genetics, 40 (5)*, p. 591-602.

Thompson, L.A., Detterman, D.K., & Plomin, R. (1991). Associations between cognitive abilities and scholastic achievement: genetic overlap but environmental differences. *Psychological Science, 2 (3)*, p. 158-165.

Trzesniewski, K.H., Moffitt, T.E., Caspi, A., Taylor, A., & Maughan, B. (2006). Revisiting the association between reading achievement and antisocial behavior: new evidence of an environmental explanation from a twin study. *Child Development, 77 (1)*, p. 72-88.

Turkheimer, E., Haley, A., Waldron, M., d'Onofrio, B., & Gottesman, I.I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science, 14 (6)*, p. 623-628.

van Beijsterveldt, C.E.M. & Van Baal, G.C.M. (2002). Twin and family studies of the human electroencephalogram: a review and a meta-analysis. *Biological Psychology, 61 (1)*, p. 111-138.

van den Oord, E., Boomsma, D., & Verhulst, F. (2000). A study of genetic and environmental effects on the co-occurrence of problem behaviors in three-year-old-twins. *Journal of Abnormal Psychology, 109 (3)*, p. 360-372.

Verweij, K.J., Zietsch, B.P., Lynskey, M.T., Medland, S.E., Neale, M.C., Martin, N.G., Boomsma, D.I., & Vink, J.M. (2010). Genetic and environmental influences on cannabis use initiation and problematic use: a meta-analysis of twin studies. *Addiction, 105 (3)*, p. 417-430.

Vinkhuyzen, A.A.E., Van Der Sluis, S., De Geus, E.J.C., Boomsma, D.I., & Posthuma, D. (2010). Genetic influences on 'environmental' factors. *Genes, Brain and Behavior, 9 (3)*, p. 276-287.

Wadsworth, S.J., Olson, R.K., & Defries, J.C. (2010). Differential genetic etiology of reading difficulties as a function of IQ: an update. *Behavior Genetics, 40 (6)*, p. 751-758.

Walker, S.O., Petrill, S.A., Spinath, F.M., & Plomin, R. (2004). Nature, nurture and academic achievement: a twin study of teacher assessments of 7-year-olds. *British Journal of Educational Psychology, 74 (3)*, p. 323-342.

Ward, M.E., McMahon, G., St Pourcain, B., Evans, D.M., Rietveld, C.A., Benjamin, D.J., Koellinger, P.D., Cesarini, D., Smith, G.D., & Timpson, N.J. (2014). Genetic variation associated with differential educational attainment in adults has anticipated associations with school performance in children. *PloS One, 9 (7)*, p. e100248

Webbink, D., Posthuma, D., Boomsma, D.I., de Geus, E.J., & Visscher, P.M. (2008). Do twins have lower cognitive ability than singletons? *Intelligence, 36 (6)*, p. 539-547.

Willcutt, E.G., Pennington, B.F., Duncan, L., Smith, S.D., Keenan, J.M., Wadsworth, S., DeFries, J.C., & Olson, R.K. (2010). Understanding the complex etiologies of developmental disorders: behavioral and molecular genetic approaches. *Journal of Developmental and Behavioral Pediatrics, 31 (7)*, p. 533-544.

de Zeeuw, E.L., van Beijsterveldt, C.E.M., Glasner, T.J., De Geus, E.J.C., & Boomsma, D.I. (2015). Arithmetic, reading and writing performance has a strong genetic component: A study in primary school children. In Preparation.

de Zeeuw, E.L., van Beijsterveldt, C.E., Glasner, T.J., Bartels, M., Ehli, E.A., Davies, G.E., Hudziak, J.J., Rietveld, C.A., Groen-Blokhuis, M.M., & Hottenga, J.J. (2014). Polygenic scores associated with educational attainment in adults predict educational achievement and ADHD symptoms in children. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 165 (6)*, p. 510-520.

de Zeeuw, E.L., van Beijsterveldt, C.E., de Geus, E.J., & Boomsma, D.I. (2012). Twin specific risk factors in primary school achievements. *Twin Research and Human Genetics, 15 (1)*, p. 107-115.

Zumberge, A., Baker, L.A., & Manis, F.R. (2007). Focus on words: a twin study of reading and inattention. *Behavior Genetics, 37 (2)*, p. 284-293.

6.

CHILDHOOD ODD AND ADHD BEHAVIOR: THE EFFECT OF CLASSROOM SHARING, GENDER, TEACHER GENDER AND THEIR INTERACTIONS

Based on Eveline L. de Zeeuw, Catherina E. M. van Beijsterveldt, Gitta H. Lubke, Tina J. Glasner, Eco J. C. de Geus and Dorret I. Boomsma (2014). Childhood ODD and ADHD behavior: The effect of classroom sharing, gender, teacher gender and their interactions. *Behavioral Genetics.* Accepted for Publication.

One criterion for a Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) diagnosis of Attention Deficit Hyperactivity Disorder (ADHD) and Oppositional Defiant Disorder (ODD) is that symptoms are present in at least two settings, and often teacher ratings are taken into account. The short Conners' Teacher Rating Scales - Revised (CTRS-R) is a widely used standardized instrument measuring ODD and ADHD behavior in a school setting. In the current study CTRS-R data were available for 7, 9 and 12-year-old twins from the Netherlands Twin Register. Measurement invariance (MI) across student gender and teacher gender was established for three of the four scales (Oppositional Behavior (OPP), Hyperactivity (HYP) and ADHD Index (ADHD)) of the CTRS-R. The fourth scale (ATT) showed an unacceptable model fit even without constraints on the data and revision of this scale is recommended. Gene-environment (GxE) interaction models revealed that heritability was larger for children sharing a classroom. There were some gender differences in the heritability of ODD and ADHD behavior and there was a moderating effect of teacher's gender at some of the ages. Taken together, this indicates that there was evidence for GxE interaction for classroom sharing, gender of the student and gender of the teacher.

INTRODUCTION

Attention deficit hyperactivity disorder (ADHD) is characterized by difficulties of both inattention and hyperactivity or impulsiveness that interfere with a child's daily functioning. At school, children have, for example, difficulty remaining in their seats and paying attention for a longer period of time. Oppositional defiant disorder (ODD) is characterized by hostile and defiant behavior towards figures with authority, going beyond normal childhood behavior. Children argue with their teacher and often lose their temper (American Psychiatric Association, 2000). Numerous studies have found a negative association between ADHD and educational achievement (Polderman et al., 2010) and children with ODD receive lower grades at school (Greene et al., 2002). Both children with ADHD and ODD are more likely to attend specialized schools.

The American Psychiatric Association (APA) estimates that 3 to 7 per cent of all school-aged children are diagnosed with ADHD, while estimates of the prevalence of ODD in children range from 2 to 16 per cent (American Psychiatric Association, 2000). It must be noted that more than 50 per cent of the children diagnosed with ADHD also have ODD (Angold, Costello & Erkanli, 1999; Wilens et al., 2002). In the general population, the ratio between boys and girls with ADHD is estimated to be 3:1, while the ratio is higher in a clinical population (Gaub & Carlson, 1997). A potential explanation of the discrepancy in the ratio between boys and girls on population versus clinical level is bias in the ratings of the teacher (Abikoff et al., 2002; Derks, Hudziak & Boomsma, 2007; Sciutto, Nolfi & Bluhm, 2004), because one criterion for a Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) diagnosis is that symptoms are present in at least two settings and often the evaluation of the teacher is taken into account. In a study focusing on children diagnosed with ADHD (Derks, Hudziak & Boomsma, 2007) teachers reported more disruptive behavior at school for boys than for girls, while there is no difference for mother ratings. For ODD, teachers also report higher prevalence rates in boys than girls while parents do not (Meisel et al., 2013). To further complicate matters, teacher bias may depend on the teacher's gender. An alternative explanation of the discrepancy is that the gender differences in ADHD and ODD behavior are more pronounced in the school environment, which may demand more of a child than the home environment.

When analyzing questionnaire data concerning psychiatric disorders, researchers often use sum scores to combine multiple items of a scale. A meaningful interpretation of a sum score is only possible when a scale measures the same disorder in all specified groups. A meaningful interpretation of a sum score is only possible when a scale measures the same disorder in all specified groups. Mellenbergh et al. (1989) defined measurement invariance (MI) with respect to group as an identical distribution of the observed sum score, conditional on the disorder that the test measures, across groups. The interpretation of group differences with respect to sum scores is only meaningful when the scale is MI (Slof-Op 't Landt MC et al., 2009). MI does not hold for example if boys score on average higher on some of the items than girls without actually scoring higher on the underlying disorder. In this case, a boy and girl, who have the same degree of a disorder, obtain systematically different sum scores. Group differences in the sum score will then reflect measurement bias instead of true underlying differences (Dolan, 2000; Mellenbergh, 1989; Meredith, 1993; Millsap & Yun-Tein, 2004).

Behavioral genetic studies have established that ADHD is amongst the most heritable psychiatric childhood disorders. According to a review of 20 twin studies, the mean estimate of the heritability of ADHD in children is over 75 per cent (Faraone et al., 2005). Estimates for ODD are somewhat lower with a heritability of around 50 per cent (Hudziak et al., 2005). Heritability estimates of problem behavior in primary school children vary widely between twins taught in the same classroom compared to twins with different teachers (Saudino, Ronald & Plomin, 2005). It is a general finding that twin correlations are larger when one teacher rates both children compared to when two teachers each rate one child. One hypothesis is that ratings could be biased due to the same person rating both children when twins are taught in the same classroom. Each teacher has his or her own perception on behavior, which can make children seem more similar when they have the same teacher (Kan et al., 2013; Simonoff et al., 1998). The second hypothesis is that there is GxE interaction (Eaves, 1984), which holds that the variation in the behavior of children in different classroom environments may depend on their genetic make-up. The classroom environment, teacher characteristics and peers differ when the twins do not share a classroom in primary school, and different environments might trigger different behavior depending on a child's genes. A study of internalizing and externalizing behavior in primary school children concluded that this was not the case, and that the heritability was higher in children sharing a classroom compared to children in different classrooms because of GxE interaction (Lamb et al., 2012). The question is whether this is also true for ODD and ADHD behavior and which differences between classrooms play a role.

In behavioral genetic studies, the absence of MI may have important consequences for heritability estimates. Absence of MI for an environmental factor, for example, gender of the teacher, could lead to differences in heritability estimates between groups (gene-environment (GxE) interaction). Absence of MI for student's gender may lead to what is known as scalar sex limitation, the effect of the genetic and environmental factors may, for example, be larger in boys than girls (Lubke, Dolan & Neale, 2004; Neale, Roysamb & Jacobson, 2006). The short Conners' Teacher Rating Scales - Revised (CTRS-R) is often filled out by teachers to assess ODD and ADHD behavior in a school setting (Conners et al., 1998). The scales of this instrument have been tested for MI in 7-year-old boys and girls (Derks et al., 2007), showing no evidence for measurement bias regarding the gender of the student. However, the study did not take into account possible differences between male and female teachers in the perception of ODD and ADHD behavior nor did it evaluate MI at older ages. Therefore, the first objective of this study is to determine whether the scales of the CTRS-R, measuring ODD and ADHD behavior, are measurement invariant for gender of the student as well as gender of the teacher throughout primary school. When measurement invariance holds, the second objective of this study is to focus on GxE interaction, and investigate whether classroom sharing, gender of the student and gender of the teacher moderate the heritability of teacher-rated ODD and ADHD behavior.

METHODS

PARTICIPANTS

The Netherlands Twin Register (NTR), established around 1987 by the Department of Biological Psychology at the VU University Amsterdam, registers approximately 40 per cent of all multiple births in the Netherlands. A survey about the development of the children is sent to the parents of the twins every two years until the twins are 12 years old (Boomsma et al., 2002; Boomsma et al., 2006; van Beijsterveldt et al., 2013). Since 1999, at approximately age 7, 9 and 12, when the twins attend primary school, parents are asked for their consent to approach the teacher(s) of their children with a survey. The survey sent to the primary school teachers includes items on background information of the teacher, functioning at school, educational achievement and the standardized questionnaires, the Teacher Report Form (TRF) (Achenbach, 1991) and the short version of the Conners' Teacher Ratings Scale - Revised (CTRS-R) (Conners, 2001).

Since 2001 data collection has yielded surveys with information on gender of the teacher for 9365, 8775 and 6649 7, 9 and 12-year-olds, respectively. We excluded children who had a disease or handicap that interfered severely with daily functioning (Age 7: N=97; Age 9: N=128; Age 12: N=95) or attended specialized education, special schools are available for children with extra needs (Age 7: N=109; Age 9: N=237; Age 12: N=226). Surveys were excluded if they were filled out by more than one teacher (Age 7: N=431; Age 9: N=259; Age 12: N=83), filled out by someone other than the regular teacher (Age 7: N=64; Age 9: N=68; Age 12: N=57), or if familiarity with the student was below average (Age 7: N=53; Age 9: N=62; Age 12: N=34). This resulted in a total sample for the measurement invariance analyses of 8611 surveys for 7-year-olds, 8021 surveys for 9-year-olds and 5954 surveys for 12-year-olds.

The sample for the GxE interaction analyses included complete phenotype data for most twin pairs (Age 7: N=3793; Age 9: N=3470; Age 12: N=2534). Incomplete data are due to only one of the teachers returning the survey. The sample consisted of 1208, 1102, and 762 twin pairs of opposite sex for respectively age 7, 9 and 12. For the same-sex twin pairs (Age 7: N=2585; Age 9: N=2368; Age 12: N=1772), determination of zygosity status was based on blood or DNA polymorphisms (Age 7: N=224; Age 9: N=331; Age 12: N=393) or on the basis of parental report of items on resemblance in appearance and confusion of the twins by parents and others (Age 7: N=2321; Age 9: N=1987; Age 12: N=1356). This last method established zygosity with an accuracy of approximately 93 per cent (Rietveld et al., 2000). Zygosity was unavailable for some twins and these twin pairs were excluded from the analyses (Age 7: N=40; Age 9: N=50; Age 12: N=23).

MEASUREMENTS

The short Conners' Teacher Rating Scale - Revised (CTRS-R) is a measurement instrument to asses ODD and ADHD behavior at school. Teachers had to indicate whether a child displayed a certain type of behavior currently or in the prior month. The short version of the CTRS-R consists of 28 items scored on a 4 point scale from 0 (not true or never) to 3 (completely true or very often) (Conners et al., 1998; Conners, 2001). The CTRS-R includes 4 scales measuring Oppositional Behavior (OPP: 5 items), Cognitive Problems/Inattention (ATT: 5 items), Hyperactivity (HYP: 7 items) and Attention Deficit Hyperactivity Disorder Index (ADHD: 12 items). One item is included in both the HYP and ADHD scale ('Easily excited, impulsive'). The item 'Inattentive, gets distracted easily' of the ADHD scale was excluded from the MI analyses as it was highly correlated with some of the other items, especially 'Easily distracted or difficulty maintaining attention' (Age 7: r = .812; Age 9: r = .805; Age 12: r = .789) and 'Short attention span' (Age 7: r = .777; Age 9: r = .716; Age 12: r = .745). As a consequence, the more stringent MI models did not converge due to multicolinearity when including this item. For the GxE interaction analyses, a sum score of a scale was computed when there was at most one missing item (OPP, ATT and HYP) or at most two missing items (ADHD) for a scale. Missing items were imputed by the rounded averaged item score of the scale for that child. The sum scores of the scales showed an L-shaped distribution and therefore the data were square root transformed prior to the analyses.

STATISTICALANALYSES

MEASUREMENT INVARIANCE

The factor structure of the four CTRS-R scales was investigated with exploratory factor analyses (EFA) with an Oblimin rotation. The number of latent factors was decided based on the scree plot and eigenvalues (larger than 1) of the factors. To test whether the scales of the CTRS-R were MI across student ('boy' or 'girl') gender and teacher ('male' or 'female') gender, multigroup (4 groups) confirmatory factor analyses (CFA) for ordinal item level data were carried out (Dolan, 2000; Meredith, 1993; Millsap & Yun-Tein, 2004) using Mplus Version 6.1 (Muthén & Muthén, 2010). With ordinal item level data an underlying continuously distributed liability is assumed and thresholds that categorize the disorder are estimated based on the response frequencies (Flora & Curran, 2004). Because of the low frequencies of the most extreme response categories, the highest two response categories were combined. The EFA and CFA models were fitted with the Theta parameterization and the weighted least squares with mean variance adjusted (WLSMV) estimator. Correction for dependency of the observations due to family clustering was done by the 'complex' option. This 'complex' option computes the standard errors and a chi square of model fit taking into account this dependency.

Different levels of MI were tested by constraining the model parameters step by step. The first level is configural invariance (configural MI), where the factor structure is the same across groups. Factor means are fixed to zero for identification purposes while factor variances, thresholds, loadings and residual variances of the continuous latent response variables are group specific. One of the factor loadings is constrained to be equal to 1 for scaling purposes. A stricter model is strong factorial invariance (strong MI), where differences in latent response means are the result of differences in the latent factor means. This model is tested by constraining both the factor loadings and thresholds to be equal across groups. The factor mean of the first group is fixed to zero and freely estimated in the other groups. The last model, strict factorial invariance (strict MI) implies that the differences in the latent response means reflect true differences in the latent factor means and variances. This is tested by constraining the factor loadings, thresholds and residual variances of the continuous latent response variables to be equal across all groups. The factor mean is still fixed to zero in the first group and freely estimated in the other groups (Dolan, 2000; Meredith, 1993; Millsap & Yun-Tein, 2004).

The root mean square error of approximation (RMSEA) and the comparative fit index (CFI) were chosen as indices of model fit. A RMSEA value smaller than .05 indicates a good fit as does a CFI value of .97 or higher (Schermelleh-Engel & Moosbrugger, 2003). The difference in goodness of fit between the nested MI models in chi square values between two nested models when using the WLMSV chi-square values is not distributed as a chi-square and as a consequence regular chi-square testing is not appropriate when using the WLSMV estimator (Muthén & Muthén, 2010). Instead, the 'difftest' option in Mplus can be used to obtain a correct chi-square difference test by using the derivatives of the variables from both models. Due to the large sample sizes these chi-square difference tests models might reject a model on the basis of a significant chi-square difference even though the model actually fit. Interpreting the chi-square as a goodness-of-fit index has been suggested as an alternative for using the chi-square as a formal test statistic. Since there are no absolute standards, a ratio between 2 and 3 is proposed to be indicative of, respectively a good and an acceptable model fit (Schermelleh-Engel & Moosbrugger, 2003). Therefore, a difference in chi-square of more than 3 times the difference in estimated parameters was interpreted as a worsening of the fit of the model. In addition, we looked at the parameter estimates and the magnitude of the modification indices to make reliable decisions on acceptance of MI.

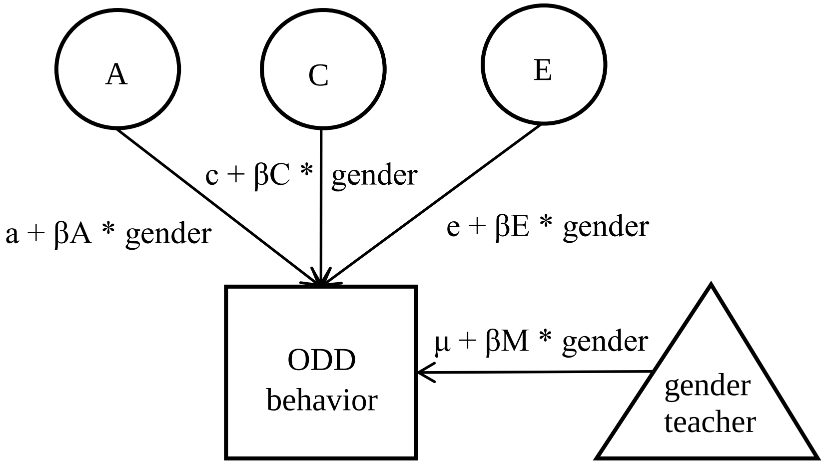
GENE-ENVIRONMENT INTERACTION MODELS

The contribution of genetic and environmental effects to the variance of the CTRS-R scales was estimated in a classical twin model (Boomsma et al., 2002; Plomin R. et al., 2008) in the R (R Core Team, 2014) package OpenMx Version 3.1.0 (Boker S.M. et al., 2011; Boker S.M. et al., 2012) with maximum likelihood estimation. First, a saturated model was fitted to the data in which means, variances and covariances were estimated in the different zygosity-by-gender groups rated by same (ST) and different (DT) teachers. Mean and variance differences between children taught by male and female, between boys and girls, between children sharing a classroom or in different classrooms and across zygosity were tested in the saturated model. It was tested whether the twin correlations could be equated between twins sharing a classroom and twins in different classrooms.

Next, GxE interaction models for gender of the student, classroom sharing and gender of the teacher were fitted to the data. GxE interaction was modelled by using multiple group designs for classroom sharing and gender of the student, and by a moderation model for teacher's gender (Figure 1) (Purcell, 2002). The models included additive genetic effects (A), dominant genetic effects (D) (or common environmental effects (C), shared by twins) and unique environmental effects (E), not shared by twins. To correct for possible confounding by gene-environment correlation (rGE), means were allowed to be different between boys and girls, between twins rated by the same or different teachers and between children rated by male or female teachers (Purcell, 2002). In the first models, differences in heritability between boys and girls were tested by constraining the estimates to be equal over gender of the student. Total variances between boys and girls were allowed to differ. Next, it was tested whether estimates could be constrained to be equal for twins rated by the same and by different teachers. Differences in genetic and environmental variance between the same and different teacher groups could be due to GxE interaction, but may also be the result of rater bias. Therefore, a correlated errors model was applied, which is an extension of the univariate twin model as it allows the unique environmental (E) effects to be correlated for twin pairs rated by the same teacher (Simonoff et al., 1998). In the last models, GxE interaction by gender of the teacher was tested by dropping from the model the moderation of the A, D (C) and E estimates by gender of the teacher.

Difference in goodness of fit of the nested models was assessed with a log-likelihood ratio test (LRT) which calculates the difference in -2log-likelihood (-2LL) between two models and evaluates this χ^2-statistic with the difference in the number of estimated parameters between the models as degrees of freedom. A p-value smaller than 0.01 was considered significant. Constraints were kept, when a more restrictive model did not significantly decrease the goodness of fit, as a more parsimonious model is preferred.

FIGURE 1 Gene-environment interaction (GxE) model with moderation by gender of the teacher



RESULTS

MEASUREMENT INVARIANCE

MI of the four scales (OPP, ATT, HYP and ADHD) of the CTRS-R was tested across gender of the student ('boy' or 'girl') and gender of the teacher ('male' or 'female') at age 7 (Age: Mean = 7.44 and SD = .47), age 9 (Age: Mean = 9.92 and SD = .53) and age 12 (Age: Mean = 12.15 and SD = .30), resulting in a 4 group comparison. Information on the gender of the teacher was available for 8611 7-year-olds (boy-male: N=322; boy-female: N=3918; girl-male: N=317; girl-female: N=4054), 8021 9-year-olds (boy-male: N=1050; boy-female: N=2841; girl-male: N=1111; girl-female: N=3019) and 5954 12-year-olds (boy-male: N=1332; boy-female: N=1503; girl-male: N=1381; girl-female: N=1738). Table 1 shows the frequencies of the item responses and the factor loadings of the items for all scales estimated from the exploratory factor analyses (EFA). Factor loadings were overall relatively high. On the basis of the scree plots and eigenvalues, a one-factor solution was chosen for OPP, ATT and HYP and a two-factor solution for ADHD (attention problems (AP) and hyperactivity/impulsivity (HI)) in all age groups (see Table 1).

Results for the tests of the three levels of MI are reported in Table S1. For OPP, HYP and ADHD the configural, strong and strict invariance models all showed an acceptable to good fit, based on the RMSEA and CFI, for all age groups. Differences in chi-square between the models with increasing equality constraints were rather small and, for the strong MI level, did not exceed more than three times the number of degrees of freedom. However, for the strict MI level, the difference in a chi-square for OPP at age 9 and HYP at age 7 and 12 was somewhat larger than this criterion, but these differences were accompanied by minor changes in RMSEA and CFI. Inspection of the modification indices revealed that they were larger for female teachers compared to male teachers for both boys and girls. Taken together, we could accept MI for the scales OPP, HYP and ADHD, for all ages, with respect to gender of the student and, more tentatively, for gender of the teacher. The fit of the MI models was acceptable to mediocre for ATT in 7-year-olds while the fit of the models was unacceptable for 9 and 12-year-olds. Even the models without constraints on the factor structure did not fit the data very well. Increasing MI levels led to a large decrease in model fit for all ages. Therefore, we could not accept MI across gender of the student and teacher for the ATT scale.

TABLE 1 Frequencies of the item responses and factor loadings as estimated in the EFA

	Age 7					Age 9					Age 12				
	Frequencies of Item Responses			Factor Loadings		Frequencies of Item Responses			Factor Loadings		Frequencies of Item Responses			Factor Loadings	
	0	1	2/3	1	2	0	1	2/3	1	2	0	1	2/3	1	2
Oppositional Behavior															
2 Defiant	.828	.141	.031	.917		.797	.167	.036	.914		.781	.181	.038	.915	
6 Defies	.901	.081	.018	.912		.894	.085	.021	.916		.876	.102	.022	.929	
10 Spiteful	.959	.034	.007	.777		.931	.059	.010	.832		.937	.054	.009	.803	
15 Argues	.862	.117	.021	.879		.841	.130	.029	.917		.817	.152	.031	..938	
20 Explosive	.921	.060	.019	.845		.907	.070	.023	.827		.916	.065	.019	.794	
Inattention/ Cognitive Problems															
4 Forgets things	.698	.225	.077	.880		.645	.260	.095	.857		.668	.248	.084	.854	
8 Poor spelling	.655	.202	.143	.881		.591	.188	.221	.860		.582	.212	.206	.862	
13 Poor reading	.696	.153	.151	.844		.728	.134	.137	.799		.786	.117	.097	.824	
18 Lacks interest	.842	.120	.039	.698		.797	.159	.045	.595		.778	.170	.052	.617	
22 Poor arithmetic	.748	.171	.081	.770		.695	.175	.130	.743		.702	.175	.123	.748	
Hyperactivity															
3 Restless	.680	.221	.099	.766		.706	.209	.085	.743		.766	.176	.058	.757	
7 Always on the go	.856	.098	.046	.830		.859	.098	.043	.794		.875	.093	.033	.794	
11 Leaves seat	.836	.115	.050	.864		.873	.090	.037	.867		.913	.066	.021	.849	
17Difficulty awaiting	.703	.204	.093	.828		.756	.167	.077	.843		.804	.140	.056	.851	
21 Runs about	.937	.047	.016	.876		.950	.038	.012	.878		.964	.028	.008	.884	
24 Difficulty playing	.776	.160	.064	.887		.788	.153	.059	.889		.826	.128	.046	.898	
27 Excitable	.798	.141	.062	.884		.799	.143	.058	.870		.826	.124	.050	.881	

ADHD Index													
Attention Problems													
14 Short attention span	.674	.214	.112	.028	.938	.687	.203	.110	.076	.897	.726	.194	.079
16 Only attention for own interests	.785	.160	.054	.202	.585	.757	.180	.063	.204	.583	.750	.184	.066
19 Distractible	.645	.231	.123	.102	.887	.649	.226	.124	.164	.832	.687	.222	.091
25 Fails to finish	.792	.164	.044	-.045	.908	.797	.163	.040	-.065	.928	.824	.142	.033
26 Not following instructions	.875	.088	.037	-.080	.925	.883	.083	.034	-.094	.949	.895	.080	.024
Hyperactivity													
5 Disturbs other children	.709	.228	.063	.855	.023	.696	.237	.067	.854	.026	.730	.210	.060
9 Cannot remain still	.779	.160	.062	.848	.106	.786	.160	.054	.808	.150	.825	.136	.039
12 Fidgets	.709	.197	.094	.676	.174	.754	.168	.078	.596	.243	.825	.132	.044
23 Interrupts	.750	.191	.059	.920	-.076	.754	.187	.059	.910	-.070	.797	.160	.043
27 Excitable	.798	.141	.062	.893	-.057	.799	.143	.058	.909	-.080	.826	.124	.050
28 Restless	.814	.129	.056	.944	.004	.821	.127	.053	.914	.035	.850	.116	.034

TABLE 2 Means and standard deviations of the untransformed sum scores of the CTRS-R scales at age 7, 9 and 12

	Male teacher								Female Teacher							
	Same Teacher				Different Teacher				Same Teacher				Different Teacher			
	Boys		Girls		Boys		Girls		Boys		Girls		Boys		Girls	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Oppositional Behavior																
Age 7	167	.8 (1.7)	170	.5 (1.3)	109	.7 (1.6)	107	.3 (1.1)	1910	.8 (1.8)	2091	.4 (1.1)	1489	.9 (1.8)	1468	.4 (1.2)
Age 9	557	1.0 (1.9)	594	.5 (1.2)	347	.9 (1.9)	349	.6 (1.6)	1401	1.1 (2.0)	1576	.5 (1.3)	1002	1.0 (2.1)	1039	.5 (1.4)
Age 12	748	1.0 (2.0)	814	.5 (1.1)	381	1.0 (1.9)	365	.6 (1.4)	805	1.2 (2.1)	959	.4 (1.2)	442	1.0 (2.2)	497	.7 (1.6)
Hyperactivity																
Age 7	167	2.5 (3.6)	170	1.5 (2.7)	108	2.3 (3.0)	106	.9 (2.1)	1907	2.7 (3.9)	2093	1.1 (2.2)	1486	2.9 (3.9)	1469	1.2 (2.3)
Age 9	556	2.3 (3.4)	592	1.0 (1.9)	347	2.3 (3.4)	351	1.1 (2.2)	1399	2.5 (3.6)	1578	.9 (1.8)	1000	2.7 (3.8)	1038	1.0 (2.3)
Age 12	752	1.8 (3.0)	815	.8 (1.8)	381	1.8 (2.8)	366	.9 (1.9)	804	2.0 (3.2)	959	.6 (1.5)	442	2.2 (3.6)	496	.9 (2.1)
ADHD Index																
Age 7	167	5.3 (6.6)	170	3.4 (5.4)	108	4.6 (5.0)	107	2.9 (4.4)	1906	5.3 (6.6)	2091	2.9 (4.6)	1485	6.2 (7.1)	1469	3.3 (4.9)
Age 9	553	5.1 (6.4)	589	2.9 (4.6)	348	5.5 (6.9)	351	3.1 (4.6)	139	5.6 (6.7)	1578	2.6 (4.2)	999	6.3 (7.0)	1039	3.0 (4.7)
Age 12	750	4.5 (6.0)	815	2.3 (3.7)	381	4.7 (5.6)	366	2.5 (3.9)	804	4.9 (6.2)	960	1.9 (3.6)	439	5.6 (6.9)	495	2.6 (4.3)

N = number of observations; SD = standard deviation

GENE-ENVIRONMENT INTERACTION MODELS

The results of the variance differences were added to the results section and the paragraph was restructured to improve clarity. 'Table 2 gives the means and standard deviations of the measurement invariant CTRS-R scales for boys and girls with the same or different male or female teachers across the three age groups. The saturated models were used to test for mean and variance differences across these groups. For OPP, there were mean and variance differences between boys and girls at all ages and variance differences across zygosity at age 7, between children sharing a classroom and children in different classrooms at age 12 and between children with the same or different male or female teachers at age 12. For HYP, there were mean and variance differences between boys and girls at all ages, mean differences across zygosity and between children sharing a classroom and children in different classrooms at age 7 and variance differences between children sharing a classroom and children in different classrooms at age 12. For ADHD, there were mean and variance differences between boys and girls at all ages and mean differences between children sharing a classroom and children in different classrooms at all ages.

TABLE 3 Twin correlations for the CTRS-R scales rated by the same teacher or different teachers at age 7, 9 and 12

	Oppositional Behavior		Hyperactivity		ADHD Index	
	ST	DT	ST	DT	ST	DT
Age 7						
MZm	.772	.495	.842	.479	.820	.555
DZm	.360	.280	.347	.289	.437	.292
MZf	.617	.394	.749	.492	.770	.514
DZf	.404	.233	.310	.211	.342	.217
DOS	.294	.112	.301	.176	.339	.250
Age 9						
MZm	.763	.334	.790	.465	.792	.447
DZm	.405	.211	.342	.208	.353	.296
MZf	.635	.442	.712	.407	.793	.497
DZf	.498	.081	.302	.145	.379	.270
DOS	.244	.133	.296	.242	.327	.254
Age 12						
MZm	.719	.518	.792	.434	.818	.546
DZm	.350	.282	.297	.310	.283	.301
MZf	.606	.500	.681	.361	.751	.414
DZf	.338	.297	.315	.282	.276	.245
DOS	.232	.185	.234	.205	.265	.233

ST = same teacher; DT = different teacher; MZm = monozygotic boys; DZm = dizygotic boys; MZf = monozygotic girls; DZf = dizygotic girls; DOS = dizygotic of opposite sex

Twin correlations for each gender by zygosity group rated by the same teacher or by different teachers are given in Table 3. For all scales, MZ correlations were higher, sometimes more than twice as high, than DZ correlations, suggesting additive (and in some cases dominant) genetic effects. Only for the OPP scale were DZ correlations larger than half the MZ correlations, suggesting common environmental effects. The GxE interaction model fitting results are reported in the online supplementary materials for the OPP (Table S2), HYP (Table S3) and ADHD (Table S4) scales of the CTRS-R. The standardized estimates (Table 4) and the contribution of the variance components (Figure 2) are given for the most parsimonious and best fitting models.

CLASSROOM SHARING

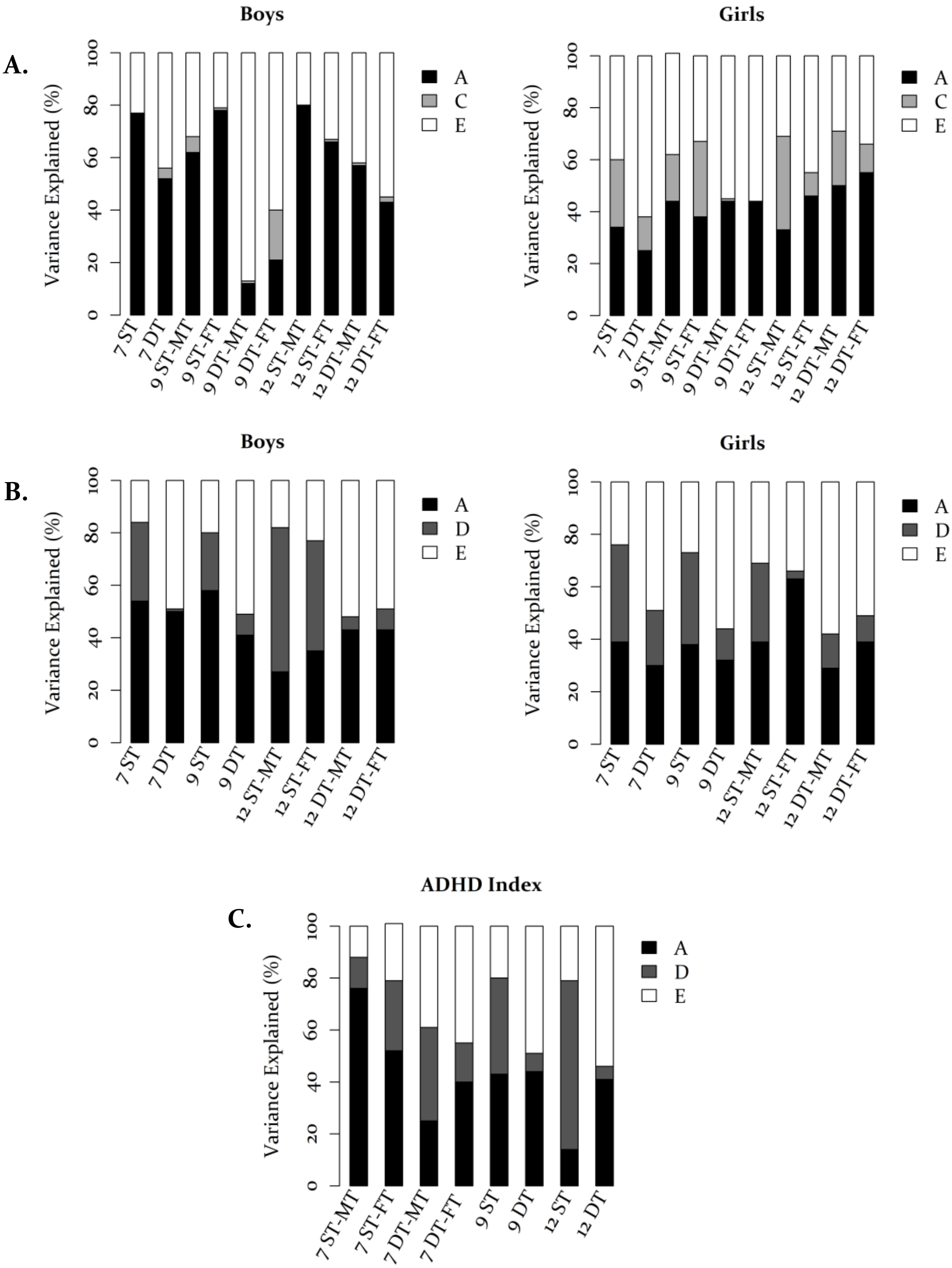
Correlations between twins rated by the same teacher could not be constrained to be equal to correlations between twins with different teachers. Constraining the variance components to be equal across same and different teachers also resulted in a significant deterioration of the model fit. A model with correlated errors was fitted to the data to check whether the differences between the same teacher and different teacher groups could be explained by rater bias. For none of the scales did the correlated errors model provide a better fit. In general, the proportion of the variance explained by genetic effects (heritability) was higher, at all ages, for children taught by the same teacher (ST) than for children rated by different teachers (DT) for OPP in boys (ST: 62-80%; DT: 12-57%) and girls (ST: 33-46%; DT: 25-55%), HYP in boys (ST: 76-84%; DT: 48-51%) and girls (ST: 66-75%; DT: 43-51%) and ADHD (ST: 78-88%; DT: 46-61%).

TABLE 4 Standardized estimates [95% Confidence intervals] of the total genetic (G), additive genetic (A), dominant genetic (D), common environmental (C) and unique environmental (E) effects on the four CTRS-R scales for 7, 9 and 12-year-olds in the best-fitting models

			Same Teacher				Different Teacher			
			G	A	C/D	E	G	A	C/D	E
Oppositional Behavior										
Age 7		Boys	.77 [.71-.80]	.77 [.71-.80]	.00 [.00-.05]	.23 [.20-.27]	.52 [.32-.62]	.52 [.32-.62]	.04 [.00-.20]	.45 [.37-.53]
		Girls	.34 [.15-.54]	.34 [.15-.54]	.26 [.07-.43]	.40 [.35-.46]	.25 [.00-.42]	.25 [.00-.42]	.13 [.00-.34]	.62 [.53-.72]
Age 9	Male	Boys	.62 [.41-.74]	.62 [.41-.74]	.06 [.00-.25]	.32 [.24-.41]	.12 [.00-.29]	.12 [.00-.29]	.01 [.00-.13]	.87 [.71-.99]
	Teacher	Girls	.44 [.20-.66]	.44 [.20-.66]	.18 [.00-.38]	.39 [.30-.50]	.44 [.30-.56]	.44 [.30-.56]	.01 [.00-.09]	.56 [.43-.70]
	Female	Boys	.78 [.69-.82]	.78 [.69-.82]	.01 [.00-.09]	.21 [.18-.25]	.21 [.00-.25]	.21 [.00-.25]	.19 [.04-.36]	.60 [.49-.72]
	Teacher	Girls	.38 [.21-.55]	.38 [.21-.55]	.29 [.13-.44]	.34 [.29-.29]	.44 [.31-.54]	.44 [.31-.54]	.00 [.00-.08]	.56 [.46-.69]
Age 12	Male	Boys	.80 [.72-.84]	.80 [.72-.84]	.00 [.00-.07]	.20 [.16-.25]	.57 [.34-.69]	.57 [.34-.69]	.01 [.00-.18]	.42 [.31-.55]
	Teacher	Girls	.33 [.13-.56]	.33 [.13-.56]	.36 [.15-.54]	.31 [.25-.38]	.50 [.27-.70]	.50 [.27-.70]	.21 [.04-.43]	.29 [.19-.42]
	Female	Boys	.66 [.53-.73]	.66 [.53-.73]	.01 [.00-.20]	.33 [.26-.41]	.43 [.22-.55]	.43 [.22-.55]	.02 [.00-.20]	.55 [.44-.68]
	Teacher	Girls	.46 [.27-.60]	.46 [.27-.60]	.09 [.00-.25]	.44 [.36-.54]	.55 [.35-.69]	.55 [.35-.69]	.11 [.00-.30]	.34 [.25-.46]
Hyperactivity										
Age 7		Boys	.84 [.81-.86]	.54 [.20-.86]	.30 [.00-.64]	.16 [.14-.19]	.51 [.43-.59]	.50 [.17-.59]	.01 [.00-.36]	.49 [.41-.57]
		Girls	.75 [.72-.78]	.39 [.16-.78]	.37 [.00-.60]	.25 [.21-.28]	.51 [.42-.58]	.30 [.04-.55]	.21 [.00-.49]	.49 [.42-.58]
Age 9		Boys	.80 [.76-.83]	.58 [.20-.82]	.22 [.00-.60]	.20 [.17-.24]	.49 [.39-.58]	.41 [.10-.57]	.08 [.00-.42]	.51 [.42-.61]
		Girls	.72 [.68-.76]	.38 [.15-.75]	.35 [.00-.58]	.28 [.24-.32]	.44 [.33-.53]	.32 [.06-.51]	.12 [.00-.41]	.56 [.47-.67]
Age 12	Male	Boys	.82 [.77-.86]	.27 [.00-.62]	.55 [.20-.82]	.18 [.14-.23]	.48 [.34-.60]	.43 [.09-.59]	.05 [.00-.40]	.52 [.40-.66]
	Teacher	Girls	.68 [.61-.74]	.39 [.00-.70]	.30 [.00-.69]	.32 [.26-.39]	.43 [.24-.57]	.29 [.03-.50]	.13 [.00-.43]	.57 [.43-.76]
	Female	Boys	.76 [.70-.81]	.35 [.03-.68]	.42 [.09-.74]	.24 [.19-.30]	.51 [.37-.62]	.43 [.11-.60]	.08 [.00-.41]	.49 [.38-.63]
	Teacher	Girls	.66 [.58-.73]	.63 [.34-.72]	.03 [.00-.33]	.34 [.27-.42]	.49 [.27-.42]	.39 [.11-.58]	.10 [.00-.38]	.51 [.39-.65]
ADHD Index										
Age 7	Male Teacher		.88 [.83-.92]	.76 [.35-.90]	.12 [.00-.53]	.12 [.08-.17]	.61 [.48-.72]	.25 [.07-.63]	.36 [.00-.62]	.39 [.28-.52]
	Female Teacher		.78 [.76-.81]	.52 [.34-.69]	.27 [.10-.44]	.22 [.19-.24]	.55 [.50-.60]	.40 [.18-.56]	.15 [.00-.38]	.45 [.40-.50]
Age 9			.80 [.77-.82]	.43 [.25-.60]	.37 [.20-.55]	.20 [.18-.23]	.50 [.43-.57]	.44 [.21-.55]	.07 [.00-.31]	.50 [.43-.57]
Age 12			.79 [.76-.81]	.14 [.00-.34]	.65 [.44-.80]	.21 [.19-.24]	.46 [.38-.53]	.41 [.11-.52]	.05 [.00-.38]	.54 [.46-.62]

G = genetic effects (summation of additive and dominant genetic effects); A = additive genetic effects; C = common environmental effects; E = unique environmental effects; MT = male teacher; FT = female teacher

FIGURE 2 The relative contribution of the additive genetic, dominant genetic, common environmental and unique environmental effects for the most parsimonious and best fitting models for Oppositional Behavior (A), Hyperactivity (B) and Attention Deficit Hyperactivity Disorder Index (C)



ST = different teacher; ST = same teacher; FT = female teacher; MT = male teacher

GENDER OF THE STUDENT

For the scales OPP and HYP, the contribution of the variance components differed between boys and girls at all ages, while this was not the case for the ADHD scale. Heritability of OPP was higher for boys (ST: 62-80%; DT: 12-57%) than girls (ST: 33-46%; DT: 25-55%). The influence of common environmental effects was, at most ages, negligible in boys (ST: 0-6%: DT: 1-19%) while it had some influence in girls (ST: 9-36%; DT: 0-21%). Heritability of HYP was slightly higher for boys (ST: 76-84%; DT: 48-51%) than girls (ST: 66-75%; DT: 43-51%). Differences between boys and girls on this scale could mainly be attributed to differences in the influence of dominant genetic effects.

GENDER OF THE TEACHER

Moderation by gender of the teacher was significant for OPP at age 9 and 12, HYP at age 12 and ADHD at age 7. For OPP at age 9, the relative influence of genetic effects was larger in boys with female teachers (ST: 78%; DT: 21%) than with male teachers (ST: 62%; DT: 12%) while it was somewhat larger for girls with male teachers (ST: 44%; DT: 44%) compared to with female teachers (ST: 38%; DT: 44%). For OPP at age 12, the opposite was true; heritability was larger in boys with male teachers (ST: 80%; DT: 57%) than with female teachers (ST: 66%; DT: 43%) while heritability was somewhat larger when girls were taught by a female teacher (ST: 46%: DT 55%) compared to when they were taught by a male teacher (ST: 33%; DT: 50%). For HYP at age 12, heritability was almost equal in boys and girls with male and female teachers, but the extent to which dominant genetic effects played a role differed across gender of the teacher. For ADHD at age 7, heritability was larger for children with male teachers (ST: 88%; DT: 61%) compared to with female teachers (ST: 78%; DT: 55%).

DISCUSSION

Three (Oppositional Behavior (OPP), Hyperactivity (HYP) and Attention Deficit Hyperactivity Disorder Index (ADHD)) of the four scales of the short Conners' Teacher Ratings Scale - Revised (CTRS-R) (Conners, 2001), used in a school setting to assess ODD and ADHD behavior, were measurement invariant across gender of the student and teacher. This means that gender differences in means and variances may be interpreted as reflecting true differences on the underlying disorder. In contrast, measurement invariance did not hold for the Inattention/Cognitive Problems (ATT) scale. Explanations for the absence of measurement invariance could be the low factor loadings and the moderate test-retest reliability of this scale. Problems with the item content have been previously suggested (Conners et al., 1998). In our sample, the internal reliability of the Inattention/Cognitive Problems scale of the short CTRS-R ranged from .78 to .82. The results of the measurement invariance analyses strongly question the reliability of this scale and its use in clinical practice. Revision of this scale is recommended as the ratings might reflect a bias instead of true differences.

Heritability of ODD and ADHD behavior, measured with the Oppositional Behavior (OPP), Hyperactivity (HYP) and Attention Deficit Hyperactivity Disorder Index (ADHD) scales of the CTRS-R is substantial. Common environmental effects had some influence on ODD behavior while dominant genetic effects had an influence on ADHD behavior. The finding of common environmental effects is consistent with earlier studies of ODD behavior using parental ratings (Burt et al., 2001; Tuvblad et al., 2009). The influence is larger in girls which may be explained by the fact that girls appear to be more sensitive to reprimands from the teacher than boys. Earlier research already concluded that girls more often feel the pressure from peers or others to behave prosocially (Roberts & Strayer, 1996). Girls might be more inclined to adapt their behavior when they are called upon by the teacher. In younger girls the common environment also has an influence when they do not share a classroom. Factors in the home environment that have been proposed to have an influence on ODD behavior are, for example, parental discipline and parental involvement (Frick et al., 1992) and the influence of these factors could depend on the gender of a child and decrease when a child grows older. The finding of dominant genetic effects for ADHD behavior, especially in children sharing a classroom, could also be due to rater contrast effects. Only when one teacher rates both children of a twin pair can the behavior of the children be contrasted and result in negative interaction effects. A higher rating for ADHD behavior in one of the children of a twin pair could lead to a lower rating for ADHD behavior in the co-twin. However, the variance in ADHD behavior is not significantly smaller in MZ twin pairs compared to DZ twin pairs, which disconfirms the presence of this type of rater bias. This is in accordance with the results of a study looking into mother and teacher ratings of hyperactivity. A contrast effect was found for the maternal ratings while the teacher ratings did not show this form of rater bias (Simonoff et al., 1998).

Heritability estimates for ADHD behavior are comparable to those found in studies taking differences between same and different teachers into account. For example, Merwood et al., (2013) also found differences in heritability between 12-year-old children sharing a classroom (76%) and not sharing a classroom (49%). One study included only twin pairs sharing a classroom and observed a heritability of 74 per cent (Hartman et al., 2007) while another included only twins not sharing a classroom and estimated a heritability of 46 per cent

(Towers et al., 2000). GxE interaction was the most plausible explanation for internalizing and externalizing problems, assessed with the Teacher Report Form, in 7 to 12-year-old twin pairs of which approximately 60 per cent shared a classroom (Lamb et al., 2012). Other studies looking into GxE interaction for ADHD in 11 to 12-year-olds (Merwood et al., 2013), and hyperactivity in 7-year olds (Saudino, Ronald & Plomin, 2005) also observed that heritability was larger when children shared a classroom. On the other hand, a study in 7-year-olds did not observe a difference between children sharing a classroom and children in different classrooms in the heritability of ODD and ADHD behavior (Derks et al., 2007), but it could be that this study did not have enough power to detect these differences in the heritability (Derks, Dolan & Boomsma, 2004).

Studies towards the heritability of teacher-rated ODD behavior are scarce. The findings of gender differences and common environmental effects were in accordance with the results of a study by Hudziak et al. (2005) that was based on a subsample of the present study. Heritability estimates for both boys (38%) and girls (21%) were somewhat different. However, this study did not take into account whether the children were rated by the same or different teachers (Hudziak et al., 2005). In contrast with current findings, none of the heritability estimates of the maternal-rated ODD behavior differed between boys and girls (Dick et al., 2005; Tuvblad et al., 2009). The differences between parent and teacher ratings of ODD behavior could be due to the fact that children can express different behavior in the classroom than they do at home. The OPP scale of the CTRS-R takes these differences into account by including different items for the teacher survey. A study observed that, although parents rated children rather similar over time, teachers with different teaching styles rated the same children very different across grades, suggesting that behavior differed in response to different teaching styles (Vitaro, Tremblay & Gagnon, 1995). Another explanation is that teachers have highly informed views on general childhood behavior for both boys and girls and are better able to assess which behavior is normative for a child of a certain age and gender.

Heritability of ODD and ADHD behavior was larger in children who shared a classroom compared to those who did not. The correlated errors model did not provide a better explanation for the differences in correlations between children rated by the same and different teachers, excluding teacher bias as an explanation, and therefore these findings are in line with GxE interaction for classroom sharing. In general, the heritability of ODD and ADHD behavior was lower in children not sharing a classroom leading to a larger impact of the environment which suggests that different behavior is elicited by different classroom environments. The children are taught by different teachers, with different rules and teaching methods and have different peers. All these factors could contribute to differences between children. For example, how teachers handle disruptive behavior is related to the behavior of a child (Rydell & Henricsson, 2004). The unique environmental variance also contains measurement error which might be increased when different teachers rate the two children of a twin pair as rater variance ends up in the measurement error (Hoyt, 2000). An important question is which differences between classroom environments play a role. Peer problems are related to ODD and ADHD behavior (Paap et al., 2013). Genetic variance in childhood aggression is moderated by peer victimization and might also moderate the heritability of ODD and ADHD (Brendgen et al., 2008). A study towards differences between monozygotic twins in their perception of the classroom environment identified, for example, the perception of a student about the relationship with the teacher as a unique environmental factor that differed between the genetically identical twins and was linked to hyperactivity as rated by the teacher (Somersalo, Solantaus & Almqvist, 2002).

For one teacher characteristic, gender, we investigated whether it moderated genetic effects on behavior in the classroom. The expression of a child's genetic vulnerability for displaying ODD and ADHD behavior at school depended in some cases on the gender of the teacher. The direction of the difference in heritability may provide an indication for one of two hypotheses. Male teachers and female teachers could provide a different learning and classroom environment with regard to, for example, structure and rules. The bioecological model (Bronfenbrenner & Ceci, 1994) predicts that the heritability of a phenotype will be lower in an adverse environment because risk environments will prevent the amplification of underlying genetic differences between children while the diathesis-stress model suggests that heritability will be higher in an adverse environment due to the expression of a genetic vulnerability that is triggered by a risk environment (Rende & Plomin, 1992). A same-gender teacher might be seen as a supportive environment as it is suggested to have a positive influence on the behavior and educational achievement of a child (Carrington, Tymms & Merrell, 2008). According to the bioecological model, genetic variation will be higher when children are taught by a same-gender teacher while the diasthesis-stress model predicts that heritability will be lower. However, in our study, the direction of the effects of gender of the teacher was not consistent which makes interpreting the GxE interaction findings difficult.

To summarize, three of the four scales of the short CTRS-R measuring teacher-rated ODD and ADHD behavior in 7, 9 and 12-year-olds were measurement invariant for student gender and teacher gender. Revision of the fourth scale (ATT) is highly recommended in order to be useable in clinical practice. The heritability of ODD and ADHD behavior was lower for children in different classrooms compared to children sharing a classroom, suggesting that different behavior is elicited by different classroom environments. Apparently, teachers, the classroom and/or peers are important environmental factors that influence the expression of ODD and ADHD behavior in primary school. The direction of the moderation of the heritability of ODD and ADHD behavior by gender of the teacher was not consistent, which makes interpretation difficult. Finding environmental factors with a moderating influence on the heritability ODD and ADHD might help improve learning environments at school to prevent manifestation of ODD and ADHD behavior in children with an increased genetic vulnerability for these disorders.

TABLE S1 Model fitting results for measurement invariance tested in three age groups across gender of the teacher and gender of the student

		N	ep	RMSEA	χ^2	CFI	χ^2 Difference Test	df	p
Oppositional Behavior									
Age 7	EFA	8552	60	.058	150.195	.994			
	Configural	8552	60	.060	173.850	.993			
	Strong	8552	36	.034	188.452	.994	50.395	24	.001
	Strict	8552	21	.039	202.633	.994	25.778	15	.040
Age 9	EFA	7962	60	.073	215.595	.993			
	Configural	7962	60	.074	237.804	.993			
	Strong	7962	36	.044	214.997	.994	33.557	24	.093
	Strict	7962	21	.042	263.845	.993	58.267	15	<.001
Age 12	EFA	5904	60	.065	130.095	.996			
	Configural	5904	60	.065	143.429	.996			
	Strong	5904	36	.041	152.748	.996	45.131	24	.006
	Strict	5904	21	.037	180.625	.996	33.185	15	.004
Cognitive Problems/ Inattention									
Age 7	EFA	8551	60	.094	382.373	.986			
	Configural	8551	60	.091	376.516	.987			
	Strong	8551	36	.079	633.634	.979	303.322	24	<.001
	Strict	8551	21	.073	723.741	.976	126.082	15	<.001
Age 9	EFA	7963	60	.145	840.426	.956			
	Configural	7963	60	.140	799.807	.963			
	Strong	7963	36	.130	1528.966	.930	765.792	24	<.001
	Strict	7963	21	.119	1721.781	.921	250.020	15	<.001
Age 12	EFA	5904	60	.147	645.088	.956			
	Configural	5904	60	.147	660.227	.961			
	Strong	5904	36	.131	1150.344	.932	530.606	24	<.001
	Strict	5904	21	.119	1291.816	.925	166.737	15	<.001
Hyperactivity									
Age 7	EFA	8552	84	.044	242.830	.995			
	Configural	8552	84	.041	261.458	.995			
	Strong	8552	48	.035	329.143	.994	100.176	36	<.001
	Strict	8552	27	.033	383.403	.993	77.061	21	<.001

Age 9	EFA	7959	84	.043	221.821	.994			
	Configural	7959	84	.043	267.452	.994			
	Strong	7959	48	.033	288.498	.994	75.832	36	<.001
	Strict	7959	27	.031	330.362	.993	59.778	21	<.001
Age 12	EFA	5904	84	.038	134.893	.995			
	Configural	5904	84	.041	194.261	.993			
	Strong	5904	48	.029	208.982	.994	50.365	36	.056
	Strict	5904	27	.032	281.340	.992	75.149	21	<.001
ADHD Index									
Age 7	EFA	8552	136	.086	2205.268	.984			
	Configural	8552	136	.070	1983.366	.986			
	Strong	8552	82	.060	1948.785	.987	100.227	54	<.001
	Strict	8552	49	.050	1661.353	.989	47.255	33	.052
Age 9	EFA	7961	136	.082	1868.673	.985			
	Configural	7961	136	.073	1979.756	.984			
	Strong	7961	82	.063	2012.996	.984	155.307	54	<.001
	Strict	7961	49	.054	1757.824	.986	60.369	33	.003
Age 12	EFA	5904	136	.078	1270.317	.985			
	Configural	5904	136	.064	1214.061	.986			
	Strong	5904	82	.054	1201.933	.987	81.171	54	.010
	Strict	5904	49	.048	1143.753	.988	90.742	33	<.001

N = number of observations; ep = estimated parameters; RMSEA = root mean square error of approximation; X2 = chi square; CFI = comparative fit index; df = degrees of freedom; EFA = exploratory factor analysis

TABLE S2 Genetic modeling results for the oppositional behavior (OPP) scale

	ep	-2ll	df	model	χ2	Δdf	p
Age 7							
0 Saturated	52	14503.83	7379	-	-	-	-
1 Saturated: ST = DT	37	14614.79	7394	0	110.96	15	<.001
2 ACE	23	14583.00	7408	0	79.17	29	<.001
3 ACE: Boys = Girls	15	14673.59	7416	2	90.58	8	<.001
4 ACE: ST = DT	17	14656.08	7414	2	73.08	6	<.001
5 ACE: Correlated Errors	18	14592.64	7413	0	87.79	33	<.001
6 ACE: FT = MT	*17*	*14587.70*	*7414*	*2*	*4.70*	*6*	*.583*
Age 9							
0 Saturated	52	14271.56	6713	-	-	-	-
1 Saturated: ST = DT	37	14417.89	6728	0	146.33	15	<.001
2 ACE	*23*	*14302.35*	*6742*	*0*	*30.79*	*29*	*.375*
3 ACE: Boys = Girls	15	14385.60	6750	2	83.25	8	<.001
4 ACE: ST = DT	17	14428.55	6748	2	126.19	6	<.001
5 ACE: Correlated Errors	18	14349.08	6747	0	75.60	33	<.001
6 ACE: FT = MT	17	14322.82	6748	2	20.47	6	.002
Age 12							
0 Saturated	52	10447.34	4913	-	-	-	-
1 Saturated: ST = DT	37	10509.68	4928	0	62.34	15	<.001
2 ACE	*23*	*10461.64*	*4942*	*0*	*14.30*	*29*	*.990*
3 ACE: Boys = Girls	15	10538.20	4950	2	76.56	8	<.001
4 ACE: ST = DT	17	10509.94	4948	2	48.30	6	<.001
5 ACE: Correlated Errors	18	10515.73	4947	0	59.45	33	.003
6 ACE: FT = MT	17	10498.14	4948	2	36.50	6	<.001

FT = female teacher; MT = male teacher; DT = different teacher; ST = same teacher; ep = estimated parameters; df = degrees of freedom; -2ll = -2loglikelihood; A = additive genetic effects; C = common environmental effects; E = unique environmental effects

TABLE S3 Genetic modeling results for the hyperactivity (HYP) scale

	ep	**-2ll**	**df**	**model**	**χ2**	**Δdf**	p
Age 7							
0 Saturated	52	20030.50	7374	-	-	-	-
1 Saturated: ST = DT	37	20187.51	7389	0	157.01	15	<.001
2 ACE	23	20063.58	7403	0	33.08	29	.275
3 ACE: Boys = Girls	15	20102.04	7411	2	38.46	8	<.001
4 ACE: ST = DT	17	20199.68	7409	2	136.10	6	<.001
5 ACE: Correlated Errors	18	20085.22	7408	0	54.00	33	.012
6 ACE: FT = MT	*17*	*20078.37*	*4709*	*2*	*14.79*	*6*	*.022*
Age 9							
0 Saturated	52	17649.84	6709	-	-	-	-
1 Saturated: ST = DT	37	17783.59	6724	0	133.76	15	<.001
2 ACE	23	17681.08	6738	0	31.24	29	.354
3 ACE: Boys = Girls	15	17707.93	6746	2	26.84	8	.001
4 ACE: ST = DT	17	17793.64	6744	2	112.56	6	<.001
5 ACE: Correlated Errors	18	17704.15	6743	0	53.63	33	.013
6 ACE: FT = MT	*17*	*17697.01*	*6744*	*2*	*15.92*	*6*	*.014*
Age 12							
0 Saturated	52	12142.50	4917	-	-	-	-
1 Saturated: ST = DT	37	12258.51	4932	0	117.01	15	<.001
2 ACE	*23*	*12176.31*	*4946*	*0*	*33.81*	*29*	*.246*
3 ACE: Boys = Girls	15	12219.40	4954	2	43.10	8	<.001
4 ACE: ST = DT	17	12249.56	4952	2	73.26	6	<.001
5 ACE: Correlated Errors	18	12216.43	4951	0	74.35	33	<.001
6 ACE: FT = MT	17	12204.84	4952	2	28.53	6	<.001

FT = female teacher; MT = male teacher; DT = different teacher; ST = same teacher; ep = estimated parameters; df = degrees of freedom; -2ll = -2loglikelihood; A = additive genetic effects; C = common environmental effects; E = unique environmental effects

TABLE S4 Genetic modeling results for the ADHD index (ADHD) scale

	ep	-2ll	df	model	χ2	Δdf	p
Age 7							
0 Saturated	52	24482.63	7369	-	-	-	-
1 Saturated: ST = DT	37	24614.63	7384	0	132.00	15	<.001
2 ACE	23	24513.40	7398	0	30.77	29	.376
3 ACE: Boys = Girls	*15*	*24533.14*	*7406*	*2*	*19.73*	*8*	*.011*
4 ACE: ST = DT	12	24640.09	7409	3	106.95	3	<.001
5 ACE: Correlated Errors	13	24549.49	7408	3	59.55	38	.014
6 ACE: FT = MT	12	24546.27	7409	3	13.13	3	.004
Age 9							
0 Saturated	52	22137.31	6703	-	-	-	-
1 Saturated: ST = DT	37	22271.04	6718	0	133.72	15	<.001
2 ACE	23	22159.55	6732	0	22.24	29	.810
3 ACE: Boys = Girls	15	22174.92	6740	2	15.37	8	.052
4 ACE: ST = DT	12	22274.78	6743	3	99.85	3	<.001
5 ACE: Correlated Errors	13	22197.56	6742	0	60.25	38	.012
6 ACE: FT = MT	*12*	*22176.08*	*6743*	*3*	*1.15*	*3*	*.765*
Age 12							
0 Saturated	52	15589.30	4912	-	-	-	-
1 Saturated: ST = DT	37	15704.31	4927	0	115.02	15	<.001
2 ACE	23	15624.83	4941	0	35.53	29	.188
3 ACE: Boys = Girls	15	15638.42	4949	2	13.59	8	.093
4 ACE: ST = DT	12	15733.73	4952	3	95.30	3	<.001
5 ACE: Correlated Errors	13	15679.13	4951	0	89.60	38	<.001
6 ACE: FT = MT	*12*	*15645.36*	*4952*	*3*	*6.94*	*3*	*.074*

FT = female teacher; MT = male teacher; DT = different teacher; ST = same teacher; ep = estimated parameters; df = degrees of freedom; -2ll = -2loglikelihood; A = additive genetic effects; C = common environmental effects; E = unique environmental effects

REFERENCES

Abikoff, H.B., Jensen, P.S., Arnold, L.L., Hoza, B., Hechtman, L., Pollack, S., Martin, D., Alvir, J., March, J.S., Hinshaw, S., Vitiello, B., Newcorn, J., Greiner, A., Cantwell, D.P., Conners, C.K., Elliott, G., Greenhill, L.L., Kraemer, H., Pelham, W.E., Jr., Severe, J.B., Swanson, J.M., Wells, K., & Wigal, T. (2002). Observed classroom behavior of children with ADHD: relationship to gender and comorbidity. *Journal of Abnormal Child Psychology, 30 (4),* p. 349-359.

Achenbach, T.M. (1991). Manual for the Child Behavior Checklist/4 - 18 and Profile. Burlington, VT: University of Vermont Department of Psychiatry.

American Psychiatric Association (2000). Diagnostic and Statistical Manual of Mental Disorders: 4th ed., text rev. DSM-IV-TR. Washington, DC: American Psychiatric Association.

Angold, A., Costello, E.J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry, 40 (01),* p. 57-87.

Boker S.M., Neale, M.C., Maes, H.H.M., Wilde M.J., Spiegel M., Brick T.R., Estabrook R., Bates T.C., Mehta P., von Oertzen T., Gore R.J., Hunter M.D., Hackett D.C., Karch J., & Brandmaier A. (2012). OpenMx 1.2 User guide.

Boker S.M., Neale, M.C., Maes, H.H.M., Wilde M.J., Spiegel M., Brick T.R., Spies J., Estabrook R., Kenny S., Bates T.C., Mehta P., & Fox J. (2011). An open source extended structural equation modeling framework. *Psychometrika, 76 (2),* p. 306-317.

Boomsma, D.I., de Geus, E.J., Vink, J.M., Stubbe, J.H., Distel, M.A., Hottenga, J.J., Posthuma, D., van Beijsterveldt, T.C., Hudziak, J.J., Bartels, M., & Willemsen, G. (2006). Netherlands Twin Register: from twins to twin families. *Twin Research and Human Genetics, 9 (6),* p. 849-857.

Boomsma, D.I., Vink, J.M., van Beijsterveldt, T.C., de Geus, E.J., Beem, A.L., Mulder, E.J., Derks, E.M., Riese, H., Willemsen, G.A., Bartels, M., van den, B.M., Kupper, N.H., Polderman, T.J., Posthuma, D., Rietveld, M.J., Stubbe, J.H., Knol, L.I., Stroet, T., & van Baal, G.C. (2002). Netherlands Twin Register: a focus on longitudinal research. *Twin Research, 5 (5),* p. 401-406.

Brendgen, M., Boivin, M., Vitaro, F., Girard, A., Dionne, G., & Perusse, D. (2008). Gene-environment interaction between peer victimization and child aggression. *Developmental Psychopathology, 20 (2),* p. 455-471.

Bronfenbrenner, U. & Ceci, S.J. (1994). Nature-nurture reconceptualized in developmental perspective: a bioecological model. *Psychological Reviews, 101 (4),* p. 568-586.

Burt, S.A., Krueger, R.F., McGue, M., & Iacono, W.G. (2001). Sources of covariation among attention-deficit/hyperactivity disorder, oppositional defiant disorder, and conduct disorder: the importance of shared environment. *Journal of Abnormal Psychology, 110 (4),* p. 516-525.

Carrington, B., Tymms, P., & Merrell, C. (2008). Role models, school improvement and the gender gap: do men bring out the best in boys and women the best in girls? *British Educational Research Journal, 34 (3),* p. 315-327.

Conners, C.K. (2001). Conners' rating scales - revised. New York, NY: Multi-Health Systems, Inc.

Conners, C.K., Sitarenios, G., Parker, J.D., & Epstein, J.N. (1998). Revision and restandardization of the Conners Teacher Rating Scale (CTRS-R): factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology, 26 (4)*, p. 279-291.

Derks, E.M., Dolan, C.V., & Boomsma, D.I. (2004). Effects of censoring on parameter estimates and power in genetic modeling. *Twin Research, 7 (6)*, p. 659-669.

Derks, E.M., Dolan, C.V., Hudziak, J.J., Neale, M.C., & Boomsma, D.I. (2007). Assessment and etiology of attention deficit hyperactivity disorder and oppositional defiant disorder in boys and girls. *Behavior Genetics, 37 (4)*, p. 559-566.

Derks, E.M., Hudziak, J.J., & Boomsma, D.I. (2007). Why more boys than girls with ADHD receive treatment: a study of Dutch twins. *Twin Reearch and Human Genetics, 10 (5)*, p. 765-770.

Dick, D.M., Viken, R.J., Kaprio, J., Pulkkinen, L., & Rose, R.J. (2005). Understanding the covariation among childhood externalizing symptoms: genetic and environmental influences on conduct disorder, attention deficit hyperactivity disorder, and oppositional defiant disorder symptoms. *Journal of Abnormal Child Psychology, 33 (2)*, p. 219-229.

Dolan, C.V. (2000). Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis. *Multivariate Behavioral Research, 35 (1)*, p. 21-50.

Eaves, J.L. (1984). The resolution of genotype x environment interaction in segregation analysis of nuclear families. *Genetic Epidemiology, 1,* p. 215-228.

Faraone, S.V., Perlis, R.H., Doyle, A.E., Smoller, J.W., Goralnick, J.J., Holmgren, M.A., & Sklar, P. (2005). Molecular genetics of attention-deficit/hyperactivity disorder. *Biological Psychiatry, 57 (11)*, p. 1313-1323.

Flora, D.B. & Curran, P.J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9 (4)*, p. 466-491.

Frick, P.J., Lahey, B.B., Loeber, R., Stouthamer-Loeber, M., Christ, M.A., & Hanson, K. (1992). Familial risk factors to oppositional defiant disorder and conduct disorder: parental psychopathology and maternal parenting. *Journal of Consulting and Clinical Psychology, 60 (1)*, p. 49-55.

Gaub, M. & Carlson, C.L. (1997). Gender differences in ADHD: a meta-analysis and critical review. *Journal of the American Academy of Child and Adolsecent Psychiatry, 36 (8)*, p. 1036-1045.

Greene, R.W., Biederman, J., Zerwas, S., Monuteaux, M.C., Goring, J.C., & Faraone, S.V. (2002). Psychiatric comorbidity, family dysfunction, and social impairment in referred youth with oppositional defiant disorder. *American Journal of Psychiatry, 159 (7)*, p. 1214-1224.

Hartman, C.A., Rhee, S.H., Willcutt, E.G., & Pennington, B.F. (2007). Modeling rater disagreement for ADHD: are parents or teachers biased? *Journal of Abnormal Child Psychology, 35 (4)*, p. 536-542.

Hoyt, W.T. (2000). Rater bias in psychological research: when is it a problem and what can we do about it? *Psychological Methods, 5 (1)*, p. 64-86.

Hudziak, J.J., Derks, E.M., Althoff, R.R., Copeland, W., & Boomsma, D.I. (2005). The genetic and environmental contributions to oppositional defiant behavior: a multi-informant twin study. *J.Am.Acad.Child Adolesc.Psychiatry, 44 (9)*, p. 907-914.

Kan, K.J., Dolan, C.V., Nivard, M.G., Middeldorp, C.M., van Beijsterveldt, C.E., Willemsen, G., & Boomsma, D.I. (2013). Genetic and environmental stability in attention problems across the lifespan: evidence from the Netherlands twin register. *Journal of the American Academy of Child and Adolsecent, 52 (1)*, p. 12-25.

Lamb, D.J., Middeldorp, C.M., van Beijsterveldt, C.E., & Boomsma, D.I. (2012). Gene-environment interaction in teacher-rated internalizing and externalizing problem behavior in 7- to 12-year-old twins. *Journal of Child Psychology and Psychiatry, 53 (8)*, p. 818-825.

Lubke, G.H., Dolan, C.V., & Neale, M.C. (2004). Implications of absence of measurement invariance for detecting sex limitation and genotype by environment interaction. *Twin Research, 7 (3)*, p. 292-298.

Meisel, V., Servera, M., Cardo, E., & Garcia-Banda, G. (2013). Prevalence of oppositional defiant disorder in a sample of Spanish schoolchildren. *Spanish Journal of Psychology, 16* p. E63.

Mellenbergh, G.J. (1989). Item bias and item response theory. *Educational Research, 13,* p. 127-143.

Meredith, W. (1993). Measurement invariance, factor-analysis and factorial invariance. *Psychometrika, 58 (4)*, p. 525-543.

Merwood, A., Greven, C.U., Price, T.S., Rijsdijk, F., Kuntsi, J., McLoughlin, G., Larsson, H., & Asherson, P.J. (2013). Different heritabilities but shared etiological influences for parent, teacher and self-ratings of ADHD symptoms: an adolescent twin study. *Psychological Medicine, 43 (9)*, p. 1973-1984.

Millsap, R.E. & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research, 39 (3)*, p. 479-515.

Muthén, L.K. & Muthén, B.O. (2010). Mplus user's guide. *6th Edition*. Los Angeles, CA: Muthén L.K. & Muthén.

Neale, M.C., Roysamb, E., & Jacobson, K. (2006). Multivariate genetic analysis of sex limitation and G x E interaction. *Twin Research and Human Genetics, 9 (4)*, p. 481-489.

Paap, M.C., Haraldsen, I.R., Breivik, K., Butcher, P.R., Hellem, F.M., & Stormark, K.M. (2013). The link between peer relations, prosocial behavior, and ODD/ADHD symptoms in 7-9-year-old children. *Psychiatry Journal, 2013,* p. 319874.

Plomin R., DeFries J.C., McClearn G.E., & McGuffin P.s (2008). Behavioral Genetics. *5thEdition.* New York, NY: Worth Publishers.

Polderman, T.J.C., Boomsma, D.I., Bartels, M., Verhulst, F.C., & Huizink, A.C. (2010). A systematic review of prospective studies on attention problems and academic achievement. *Acta Psychiatrica Scandinavica, 122 (4)*, p. 271-284.

Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research, 5 (6)*, p. 554-571.

R Core Team (2014). R: A language and environment for statistical computing. Vieanna, Austria: R Foundation for Statistical Computing.

Rende, R. & Plomin, R. (1992). Diathesis-stress models of psychopathology: A quantitative genetic perspective. *Applied and Preventive Psychology, 1 (4)*, p. 177-182.

Rietveld, M.J., van, D., V, Bongers, I.L., Stroet, T.M., Slagboom, P.E., & Boomsma, D.I. (2000). Zygosity diagnosis in young twins by parental report. *Twin Research, 3 (3)*, p. 134-141.

Roberts, W. & Strayer, J. (1996). Empathy, emotional expressiveness, and prosocial behavior. *Child Development, 67 (2)*, p. 449-470.

Rydell, A.M. & Henricsson, L. (2004). Elementary school teachers' strategies to handle externalizing classroom behavior: a study of relations between perceived control, teacher orientation and strategy preferences. *Scandinavian Journal of Psychology, 45 (2)*, p. 93-102.

Saudino, K.J., Ronald, A., & Plomin, R. (2005). The etiology of behavior problems in 7-year-old twins: substantial genetic influence and negligible shared environmental influence for parent ratings and ratings by same and different teachers. *Journal of Abnormal Child Psychology, 33 (1)*, p. 113-130.

Schermelleh-Engel, K. & Moosbrugger, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8,* p. 23-74.

Sciutto, M.J., Nolfi, C.J., & Bluhm, C. (2004). Effects of child gender and symptom type on referrals for ADHD by elementary school teachers. *Journal of Emotional and Behavioral Disorders, 12 (4)*, p. 247-253.

Simonoff, E., Pickles, A., Hervas, A., Silberg, J.L., Rutter, M., & Eaves, L. (1998). Genetic influences on childhood hyperactivity: contrast effects imply parental rating bias, not sibling interaction. *Psychological Medicine, 28 (4)*, p. 825-837.

Slof-Op 't Landt MC, Dolan, C.V., Rebollo-Mesa, I., Bartels, M., Van Furth, E.F., van Beijsterveldt, C.E., Meulenbelt, I., Slagboom, P.E., & Boomsma, D.I. (2009). Sex differences in sum scores may be hard to interpret: the importance of measurement invariance. *Assessment, 16 (4)*, p. 415-423.

Somersalo, H., Solantaus, T., & Almqvist, F. (2002). Classroom climate and the mental health of primary school children. *Nordic Journal of Psychiatry, 56 (4)*, p. 285-290.

Towers, H., Spotts, E., Hetherington, E.M., Plomin, R., & Reiss, D. (2000). Genetic and environmental influences on teacher ratings of the Child Behavior Checklist. *International Journal of Behavioral Development, 24,* p. 373-381.

Tuvblad, C., Zheng, M., Raine, A., & Baker, L.A. (2009). A common genetic factor explains the covariation among ADHD ODD and CD symptoms in 9-10 year old boys and girls. *Journal of Abnormal Child Psychology, 37 (2)*, p. 153-167.

van Beijsterveldt, C.E., Groen-Blokhuis, M., Hottenga, J.J., Franic, S., Hudziak, J.J., Lamb, D., Huppertz, C., de Zeeuw, E., Nivard, M., Schutte, N., Swagerman, S., Glasner, T., van Fulpen, M., Brouwer, C., Stroet, T., Nowotny, D., Ehli, E.A., Davies, G.E., Scheet, P., Orlebeke, J.F., Kan, K.J., Smit, D., Dolan, C.V., Middeldorp, C.M., de Geus, E.J., Bartels, M., & Boomsma, D.I. (2013). The Young Netherlands Twin Register (YNTR): longitudinal twin and family studies in over 70,000 children. *Twin Research and Human Genetics, 16 (1)*, p. 252-267.

Vitaro, F., Tremblay, R.E., & Gagnon, C. (1995). Teacher ratings of children's behaviors and teachers' management styles: a research note. *Journal of Child Psychology and Psychiatry, 36 (5)*, p. 887-898.

Wilens, T.E., Biederman, J., Brown, S., Tanguay, S., Monuteaux, M.C., Blake, C., & Spencer, T.J. (2002). Psychiatric comorbidity and functioning in clinically referred preschool children and school-age youths with ADHD. *Journal of the American Academy of Child & Adolescent Psychiatry, 41 (3)*, p. 262-268.

Part III

MOLECULAR GENETICS AND CAUSALITY

7.

POLYGENIC SCORES ASSOCIATED WITH EDUCATIONAL ATTAINMENT IN ADULTS PREDICT EDUCATIONAL ACHIEVEMENT AND ADHD SYMPTOMS IN CHILDREN

Based on de Zeeuw, EL, van Beijsterveldt, CE, Glasner, TJ, Bartels, M, Ehli, EA, Davies, GE, Hudziak, JJ, Consortium SSGA, Rietveld, CA, Groen-Blokhuis, MM, Hottenga, JJ, de Geus, EJ, & Boomsma, DI (2014). Polygenic scores associated with educational attainment in adults predict educational achievement and ADHD symptoms in children. *American Journal of Medical Genetics Part B, 165 (6),* p. 1-11.

The American Psychiatric Association estimates that 3 to 7 per cent of all school aged children are diagnosed with attention deficit hyperactivity disorder (ADHD). Even after correcting for general cognitive ability, numerous studies report a negative association between ADHD and educational achievement. With polygenic scores we examined whether genetic variants that have a positive influence on educational attainment have a protective effect against ADHD. The effect sizes from a large GWA meta-analysis of educational attainment in adults were used to calculate polygenic scores in an independent sample of 12-year-old children from the Netherlands Twin Register. Linear mixed models showed that the polygenic scores significantly predicted educational achievement, school performance, ADHD symptoms and attention problems in children. These results confirm the genetic overlap between ADHD and educational achievement, indicating that one way to gain insight into genetic variants responsible for variation in ADHD is to include data on educational achievement, which are available at a larger scale.

INTRODUCTION

The American Psychiatric Association (APA) estimates that 3 to 7 per cent of all school aged children are diagnosed with attention deficit hyperactivity disorder (ADHD) (American Psychiatric Association,2000). These children have difficulties with inattention, impulsivity and/or hyperactivity interfering with their daily functioning. Children with ADHD have, on average, lower general cognitive ability than controls (Frazier et al.,2007). At school, they are more likely to repeat a grade and are more often referred to specialized education (Biederman et al.,1996). In addition, lower educational attainment is negatively related to numerous outcomes, including earnings (Julian T. and Kominski R.,2011) and health (Mackenbach et al.,1997). The phenotypic association between ADHD and general cognitive ability and ADHD and educational achievement is estimated to be around -0.30 (Kuntsi et al.,2004; Polderman et al.,2010). This correlation is also shown longitudinally; attention problems assessed at an earlier age predicted lower general cognitive ability and educational achievement years later (Polderman et al.,2006; Polderman et al.,2010). Twin and family studies showed that a substantial part of the (longitudinal) associations seems to be due to shared genetic influences (Kuntsi et al.,2004; Polderman et al.,2006; Saudino and Plomin,2007). However, there are no studies that examined at the genotype level whether ADHD and educational achievement share the same common genetic variants.

ADHD in children is approximately 75 per cent heritable (Faraone et al.,2005) and several candidate genes have been identified (Mick et al.,2010; Neale et al.,2010a; Neale et al.,2010b; Stergiakouli et al.,2012), but the largest meta-analysis of genome-wide association (GWA) studies with 2,064 trios, 896 cases and 2,455 controls has not led to the discovery of causal genetic variants associated with ADHD (Neale et al.,2008).One of the explanations is that ADHD is a highly complex disorder caused by many common genetic variants with small effects. The non-significant single nucleotide polymorphisms (SNPs) probably captured relevant genetic variation, but sample sizes have not been large enough to detect these small effects (Flint and Munafo,2013; Neale et al.,2010b). The same is true for educational achievement. It is a trait which is heritable, with heritability estimates of 60–70 per cent in children in current Western society (Bartels et al.,2002; Kovas et al.,2007), with a substantial phenotypic and genetic correlation with general cognitive ability (Bartels et al.,2002), and approximately 40 per cent in adults (Rietveld et al.,2013). A large GWA study of educational attainment (126,559 adult individuals) revealed genome-wide significant associated genetic variants with a largest estimated effect of 0.02 per cent (Rietveld et al.,2013). In an additional analysis, (Rietveld et al.,2013) combined the effect of all genetic variants, including non-significant variants, and explained approximately 20 per cent of the variance in educational attainment, indicating that educational attainment too is a very polygenic phenotype.

By using polygenic scores the information from non-significant genetic variants can be used to test whether these genetic variants with small effects may actually explain a part of the variance (Purcell et al.,2009). Polygenic scores also allow for exploration of the underlying etiology of the association between two phenotypes, such as, ADHD and educational achievement. Here, it is expected that genetic variants associated with one phenotype, will explain part of the variance in the other phenotype. Recently, this method was applied by (Lencz et al.,2014), who compared polygenic scores, consisting of genetic variants related to general cognitive ability, between schizophrenic patients and controls, showing that the schizophrenic group had lower polygenic scores than the control group, This suggests that some of the genetic variants are involved in both general cognitive ability and schizophrenia. No study has yet used polygenic scores to determine whether the same common genetic variants underlie both ADHD and educational achievement.

The expression of the genotype of an individual may depend on age, with different genes influencing a phenotype in adults and children. General cognitive ability becomes more heritable over time although the influencing genes remain relatively stable (Deary et al.,2012; Franic et al.,2014; Haworth et al.,2010). In childhood, there is a large overlap between genes that have an influence on educational achievement at the start of primary school and genes that are influencing the trait at the end of primary school (Kovas et al.,2007). However, less is known about the genetic stability of educational achievement from childhood into adulthood. Therefore, we first determined whether the same genetic variants contribute to educational attainment in adults and educational achievement and school performance in children. In the present study, polygenic scores, based on the effect sizes from the meta-analysis of educational attainment by (Rietveld et al.,2013), were calculated for children of primary school age and used to explore the association between educational achievement and ADHD symptoms and attention problems.

METHODS

PARTICIPANTS

The Netherlands Twin Register (NTR), established around 1987 by the Department of Biological Psychology at the VU University Amsterdam, registers approximately 40 per cent of all multiple births in the Netherlands (Boomsma et al.,2006; van Beijsterveldt et al.,2013). The parents of the twins receive a survey about the development of their children every two years until the twins are 12 years old. Since 1999, at age 7, 9 and 12, when the twins attend primary school, parents are asked for their consent for the NTR to approach the teacher(s) of their children with a survey. Genotypes and data for educational achievement, school performance, ADHD symptoms and/or attention problems at age 12 were available for 2133 children. Data were excluded if a child had a non-European ancestry (N=106) or had a disease or handicap that interfered severely with daily functioning (N=38). The ancestry was determined on the basis of a child's genotype data by using the 1000 Genomes dataset as a reference and was confirmed by the data on birth country of the parents of the child. Eight principal components of the 1000 Genomes cluster the European populations together and a child was labeled non-Dutch ancestry when the child was an outlier on one of these principal components (Abdellaoui et al.,2013). This resulted in a total sample of 1,989 12-year-old children with genotype data. The children belonged to 1,030 families.

MEASUREMENTS

Educational achievement was assessed by a score on a national standardized test of educational achievement, which is administered in the last grade of primary school at approximately 80 per cent of all schools in the Netherlands (Cito,2002). The test consists of multiple choice items in four different subjects, namely Arithmetic, Language, Study Skills and Science and Social Studies. The first three subscales are combined into a Total Score, which is standardized on a scale between 500 and 550. Due to the fact that the Total Score of the educational achievement test has also been collected in other surveys send to the parents, teacher and the children themselves, there are more data available for the Total Score than for the subscales.

School performance was assessed by teacher ratings of three educational domains, namely Arithmetic, Language and Reading, with two versions of the teacher survey. In the first version (birth cohorts 1989–1993), teachers could choose up to six educational domains and rate the proficiency of the students on a five-point scale from 1 (insufficient) to 5 ((very) good). In the second version (birth cohorts 1994–2000), teachers rated the proficiency of the students in four predefined educational domains on the same five-point scale. Due to the free choice in the first version of the survey, the sample size for the teacher ratings differs across educational domains.

Attention Problems (AP) were assessed, by teachers, with the Teacher Report Form (TRF) and, by mothers, with the Child Behavioral Check List (CBCL) (Achenbach,1991). The TRF AP scale consists of 20 items rated on a 3 point scale fromo (not at all or never) to 2 (very true or often). The CBCL AP scale consists of 10 items, which overlap with the TRFAP scale. ADHD symptoms were assessed, by teachers, with the ADHD index of the short version of the Conners' Teacher Rating Scales - Revised (CTRS-R) and, by mothers, with the ADHD index of the short version of the Conners' Parent Rating Scales - Revised (CPRS-R) (Conners et al.,1998). Both scales consist of 12 items, of which 6 overlap, rated on a 4 point scale from 0 (not true or never) to 3 (completely true or very often). Sum scores were computed when subjects had no or a limited number of missing items on a scale (no more than two for ADHD symptoms and mother-rated AP and no more than three for teacher-rated AP). A missing item on a scale was imputed by taking the rounded average of the scale for that child.

GENOTYPE DATA

DISCOVERY SAMPLE

The GWA meta-analysis on educational attainment conducted in a discovery sample of 126,559 individuals is described in (Rietveld et al.,2013). Educational attainment was analyzed as a quantitative measure defined as an individual's total years of schooling (EduYears) and as a binary measure defined as whether or not an individual had a college degree (College). The GWA meta-analysis discovery sample included adult NTR participants, who could be related to the children in the sample used in the present study. Therefore, the meta-analysis was carried out again and effect sizes were estimated based on a sample without the NTR individuals.

TARGET SAMPLE

Selection of NTR samples for genotyping gave preference to samples which were collected in unselected groups of children who participated in cognition studies (Hoekstra, Bartels and Boomsma,2007; Polderman et al.,2006), in MRI projects (van Soelen et al.,2012), and in a study of blood group chimerism (van Dijk, Boomsma and de Man,1996). Approximately one third of the children took part in a study (Derks et al.,2008) that selected children for the presence or absence of high AP/ADHD symptom scores. Children with longitudinal DNA samples, or whose parents could be genotyped, also were included (Scheet et al.,2012). The remaining samples were selected based on the availability of longitudinal phenotype data collected in NTR survey studies. There were no significant differences for educational achievement, school performance, attention problems and ADHD symptoms between the genotyped group (N=1989) and the group of children without genotype data (N=16581). NTR individuals were genotyped on Illumina and Affymetrix 6.0 platforms. Data were phased using Mach 1.0 and genotype imputation was performed on a larger sample with Minimac using the 1000 Genome Phase I Integrated Release Version 3 Haplotypes (b37, HG19). For the quality control (QC) of the single nucleotide polymorphisms (SNP), criteria were a Hardy-Weinberg equilibrium (HWE) p-value >0.00001, minor allele frequency (MAF) >0.01, call rate>0.95, Mendel error rate <0.02 and allele frequency difference with reference <0.20 and, for C/G and A/T, SNPs were only included with a MAF<0.35. For the QC of the samples, criteria were a missing rate <0.10, heterozygosity -0.10<F<0.10, consistency between reported gender and sex chromosome genotypes, consistency of expected and observed family relatedness between samples and Mendelian error rate<0.02.

POLYGENIC SCORES

The SNPs were clumped using the linkage disequilibrium (LD) based result clumping procedure in Plink (Purcell et al.,2007). All SNPs were included with the default settings of a LD threshold based on a R2 of 0.50 and a distance threshold of 250 kb to ensure that only nearly independent SNPs were included in the calculation of the polygenic scores. Only SNPs overlapping between the discovery and the target sample were included in the clumping procedure. The remaining criteria for the SNPs were a MAF >0.01, an info score >0.40 in both the discovery and the target sample, a difference in the frequency of the reported minor alleles <0.15, and, for C/G and A/T SNPs, a MAF <0.35. The clumping procedure led to a total number of 343,388 and 352,397 SNPs for the calculation of the polygenic scores based on, respectively, EduYears and College. The polygenic scores were then calculated in the Plink program (Purcell et al.,2007) for each individual by multiplying the number of observed effect alleles with the effect size found in the meta-analysis and summed over all SNPs (Purcell et al.,2009). The effect size of a SNP was calculated by multiplying the METAL (Willer, Li and Abecasis,2010) z-statistic with the square root of twice the MAF times the major allele frequency (Rietveld et al.,2013). For each individual, two polygenic scores were calculated; one based on the effect sizes for EduYears and the other on those for College. The polygenic scores were calculated for different subsets of SNPs, selected on the bases of their p-value in the discovery sample, with thresholds of $5x10^{-8}$, $1x10^{-5}$, 0.001, 0.01, 0.05, 0.1, 0.5 and 1.0.

STATISTICAL ANALYSES

Educational achievement, school performance, AP and ADHD symptoms were regressed on the EduYears and College polygenic scores in linear mixed models using the Statistical Package for the Social Sciences (SPSS 20) (2011). For each analysis, the predictor and the outcome measure were standardized within each subset of children with data available on both. To correct for dependency of the observations due to family clustering an additive genetic variance component was included as a random effect based on the family pedigree and dependent on zygosity. Covariates included in the analyses were gender, three principal components, reflecting ancestry differences within the Dutch population, six principal components, reflecting ancestry differences based on the 1000 Genomes project, one principal component, correcting for a possible batch effect and a number of dummy variables to indicate the specific subset in which individuals were genotyped, to correct for differences in genotyping quality across sets (Abdellaoui et al.,2013). The amount of variance (R2) explained by the polygenic scores was calculated by squaring the standardized regression coefficient. The total number of independent dimensions in the outcome data was extracted from the correlation matrix of the phenotypes with the MatSpD program developed by (Nyholt,2004). The phenotypes contained 7 independent dimensions (Li and Ji,2005) and therefore a p-value of 0.007 (0.05/7) was considered significant.

RESULTS

First, it was tested whether the polygenic scores based on educational attainments, as obtained in an adult sample, predicted educational achievement and school performance in children. Both genotype and phenotype data were available for educational achievement in Arithmetic (*N* = 745, Mean = 59.5, SD = 27.2), Language (*N* = 746, Mean = 58.7, SD = 27.2), Study Skills (*N* = 744, Mean = 61.0, SD = 26.4), Science and Social Studies (*N* = 662, Mean = 57.5, SD = 27.8) and the Total Score (*N* = 1365, Mean = 537.7, SD = 8.4) and for school performance in Arithmetic (N = 901, Mean = 3.83, SD = 1.13), Language (N = 847, Mean = 3.84, SD = .97) and Reading (N = 470, Mean = 3.79, SD = 1.08). The standardized regression coefficients are reported in Table 1. The results are given for the polygenic scores based on years of education (EduYears) and for completion of College (College). The polygenic scores significantly predicted educational achievement and school performance at certain thresholds. All significant effects were in the expected direction. Polygenic scores, related to years of education and completion of a college degree, predicted higher scores on the educational achievement test and higher ratings for school performance. The highest proportion of variance explained by the EduYears (Figure 1) and College (Figure 2) polygenic scores was at different thresholds (p_T) for educational achievement in Arithmetic (EduYears: R^2 = .012, p = .006 at p_T = 0.5; College: R^2 = .021, p = $3x10^{-4}$ at p_T = $1x10^{-5}$), Language (EduYears: R^2 = .021, p = $4x10^{-4}$ at p_T = 1.0; College: R^2 = .028, p = $8x10^{-5}$ at p_T = 1.0), Study Skills (EduYears: R^2 = .016, p = .002 at p_T = 0.5; College: R^2 = .017, p = .002 at p_T = 1.0), Science and Social Studies (EduYears: R^2 = .006, p = .060 at p_T = 1.0; College: R^2 = .013, p = .008 at p_T = 0.1) and the Total Score (EduYears: R^2 = .024, p = $4x10^{-7}$ at p_T = 0.5; College: R^2 = .022, p = $9x10^{-7}$ at p_T = 1.0) and for school performance on Arithmetic (EduYears: R^2 = .025, p = $2x10^{-5}$ at p_T = 0.1; College: R^2 = .027, p = $1x10^{-5}$ at p_T = 0.5), Language (EduYears: R^2 = .033, p = $2x10^{-6}$ at p_T = 1.0; College: R^2 = .025, p = $4x10^{-5}$ at p_T = 0.5) and Reading (EduYears: R^2 = .031, p = $4x10^{-4}$ at p_T = 1.0; College: R^2 = .042, p = $1x10^{-4}$ at p_T = 0.05).

Both genotype and phenotype data were available for AP (N = 1028, Mean = 4.63, SD = 5.71) and ADHD index (N = 583, Mean = 3.92, SD = 5.78) rated by teachers and AP (N = 1856, Mean = 2.72, SD = 2.99) and ADHD index (N = 1164, Mean = 6.48, SD = 6.91) rated by mothers. The polygenic scores, based on EduYears and College, significantly predicted the score of the ADHD index at certain p-value thresholds (Table 1). All significant effects were in the expected direction. Higher polygenic scores were associated with a lower score on the ADHD index, especially for the larger sample of mother ratings. For AP, the effects were less clear. There was only one polygenic score, based on College at the threshold of suggestive genome-wide significant association, for AP rated by teachers. The highest proportion of variance explained by the EduYears (Figure 1) and College (Figure 2) polygenic scores was at different thresholds (p_T) for teacher-rated AP (EduYears: R^2 = .006, p = .022 at p_T = 0.1; College: R^2 = .014, p = $5x10^{-4}$ at p_T = $1x10^{-5}$) and ADHD index (EduYears: R^2 = .011, p = .016 at p_T = 0.1; College: R^2 = .021, p = .002 at p_T = 0.001) and for mother-rated AP (EduYears: R^2 = .002, p = .098 at p_T = 1.0; College: R^2 = .005, p = .010 at p_T = $1x10^{-5}$) and ADHD index (EduYears: R^2 = .014, p = $2x10^{-4}$ at p_T = .1; College: R^2 = .009, p = .003 at p_T = $1x10^{-5}$).

To investigate whether the non-normality of the school performance and ADHD phenotypes influenced the results, all analyses for school performance and attention problems and ADHD symptoms were also conducted after normalizing the data in PRELIS (Jöreskog and Sörbom,2002). PRELIS transforms the data by fitting an inverse normal density function to the ranked data. This normalization did not alter the results of the polygenic score analyses. Almost exactly the same pattern of significant regression coefficients was observed, with slightly higher estimates for the ADHD related phenotypes (results are available upon request from the first author).

TABLE 1 The standardized regression coefficients (β) for the association between the polygenic scores, based on the clumped results for years of education (EduYears) for Educational Achievement, School Performance, Attention Problems (AP) and ADHD symptoms, rated by teacher and mother

	Educational Achievement									
	Arithmetic		Language		Study Skills		Science and Social Studies		Total Score	
p-value Threshold	β	p	β	p	β	p	β	P	β	p
5x10-8	.032	.423	.003	.935	-.025	.553	-.013	.763	.019	.540
1x10-5	.042	.296	.043	.312	.041	.331	.043	.306	.051	.100
.001	.093	.020	.093	.027	.074	.078	.044	.295	.097	.002
.01	.102	.010	.147	4x10^{-4}	.110	.008	.071	.087	.127	4x10^{-5}
.05	.095	.016	.140	7x10^{-4}	.104	.012	.071	.090	.145	2x10^{-6}
.1	.095	.015	.128	.002	.105	.010	.059	.156	.154	4x10^{-7}
.5	.110	.006	.146	4x10^{-4}	.128	.002	.078	.060	.154	4x10^{-7}
1	.103	.009	.146	4x10^{-4}	.126	.002	.078	.060	.154	4x10^{-7}

	School Performance						Attention Problems				ADHD			
	Arithmetic		Language		Reading		Teacher		Mother		Teacher		Mother	
p-value Threshold	β	p	β	P	β	p	β	p	β	p	β	p	β	p
5x10^{-8}	.026	.488	-.007	.858	-.031	.532	.006	.858	-.005	.839	.061	.174	.015	.646
1x10^{-5}	.086	.023	.087	.023	.021	.682	-.120	5x10^{-4}	-.068	.010	-.117	.010	-.093	.003
.001	.059	.119	.066	.087	.118	.022	-.044	.209	-.004	.874	-.145	.002	-.091	.005
.01	.062	.099	.106	.006	.195	2x10^{-4}	-.022	.522	.007	.803	-.087	.059	-.061	.054
.05	.130	6x10^{-4}	.147	1x10^{-4}	.204	1x10^{-4}	-.043	.224	-.013	.623	-.094	.041	-.080	.012
.1	.139	3x10^{-4}	.151	9x10^{-5}	.196	2x10^{-4}	-.048	.173	-.026	.330	-.107	.020	-.093	.004
.5	.164	1x10^{-5}	.158	4x10^{-5}	.192	2x10^{-4}	-.057	.099	-.022	.402	-.099	.029	-.083	.008
1	.163	1x10^{-5}	.152	7x10^{-4}	.186	3x10^{-4}	-.061	.077	-.025	.344	-.100	.028	-.087	.006

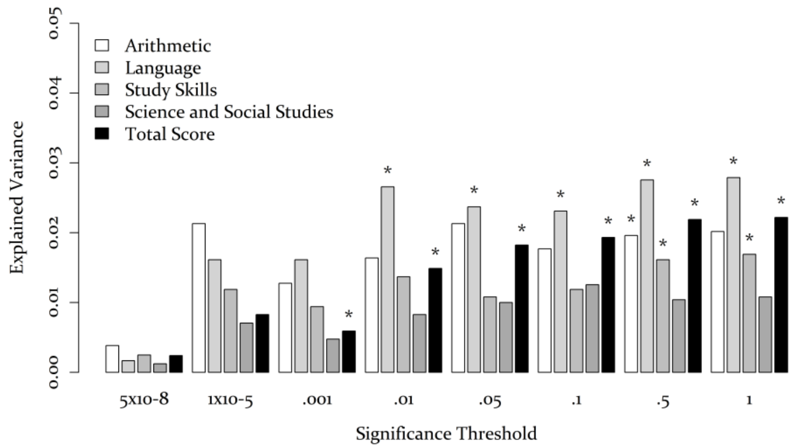
TABLE 2 The standardized regression coefficients (β) for the association between the polygenic scores, based on the clumped results for completion of a college degree (College) for Educational Achievement, School Performance, Attention Problems (AP) and ADHD symptoms, rated by teacher and mother

	Educational Achievement									
	Arithmetic		Language		Study Skills		Science and Social Studies		Total Score	
p-value Threshold	β	p	β	p	β	p	β	P	β	p
$5x10^{-8}$	.062	.124	.041	.334	.050	.238	.035	.413	.049	.108
$1x10^{-5}$	.146	$3x10^{-4}$	.127	.003	.109	.011	.084	.050	.091	.003
.001	.113	.005	.127	.003	.097	.021	.069	.102	.077	.012
.01	.128	.001	.163	$1x10^{-4}$	.117	.005	.091	.031	.122	$5x10^{-5}$
.05	.146	$3x10^{-4}$	.154	$3x10^{-4}$	.104	.014	.100	.018	.135	$8x10^{-6}$
.1	.133	.001	.152	$4x10^{-4}$	.109	.010	.112	.008	.139	$5x10^{-6}$
.5	.140	$5x10^{-4}$	.166	$8x10^{-5}$	.127	.002	.102	.016	.148	$1x10^{-6}$
1	.142	$4x10^{-4}$	.167	$8x10^{-5}$	.130	.002	.104	.013	.149	$9x10^{-7}$

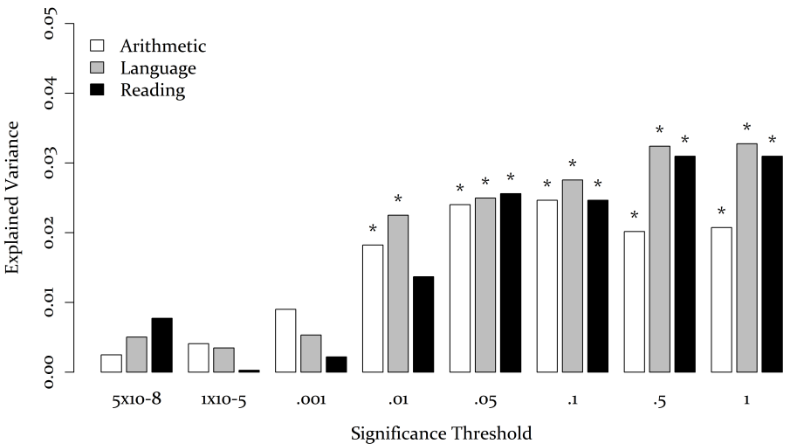
	School Performance						Attention Problems				ADHD			
	Arithmetic		Language		Reading		Teacher		Mother		Teacher		Mother	
p-value Threshold	β	p	β	P	β	p	β	p	β	p	β	p	β	p
$5x10^{-8}$	.050	.183	.071	.061	.088	.079	.028	.424	.003	.916	-.035	.449	-.024	.460
$1x10^{-5}$	.064	.085	.059	.123	.017	.738	.006	.856	-.008	.756	-.031	.495	-.066	.045
.001	.095	.011	.073	.054	.047	.362	-.047	.176	-.038	.143	-.053	.242	-.063	.046
.01	.135	$2x10^{-4}$	.150	$6x10^{-5}$	.117	.019	-.085	.013	-.044	.098	-.107	.017	-.111	$6x10^{-4}$
.05	.155	$2x10^{-5}$	.158	$2x10^{-5}$	.160	.001	-.061	.075	-.035	.186	-.091	.040	-.108	$9x10^{-4}$
.1	.157	$2x10^{-5}$	.166	$8x10^{-6}$	.158	.001	-.078	.022	-.040	.136	-.107	.016	-.119	$2x10^{-4}$
.5	.142	$1x10^{-4}$	.180	$2x10^{-6}$	.176	$4x10^{-4}$	-.078	.023	-.044	.098	-.100	.027	-.099	.002
1	.144	$9x10^{-5}$	.181	$2x10^{-6}$	.176	$4x10^{-4}$	-.077	.026	-.043	.107	-.096	.033	-.099	.002

FIGURE 1 The variance explained (R^2) in Educational Achievement (A), School Performance (B) and Attention Problems (AP)/ADHD symptoms (C) by the polygenic scores based on the clumped results for years of education (EduYears) calculated with different p-value thresholds

A.



B.



C.

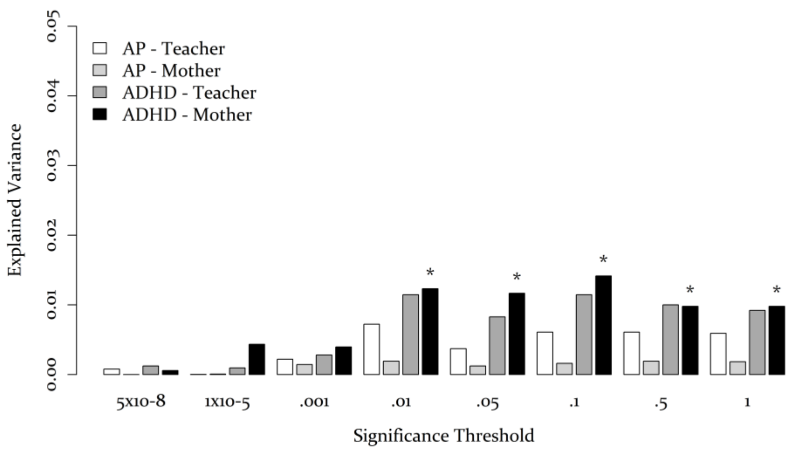
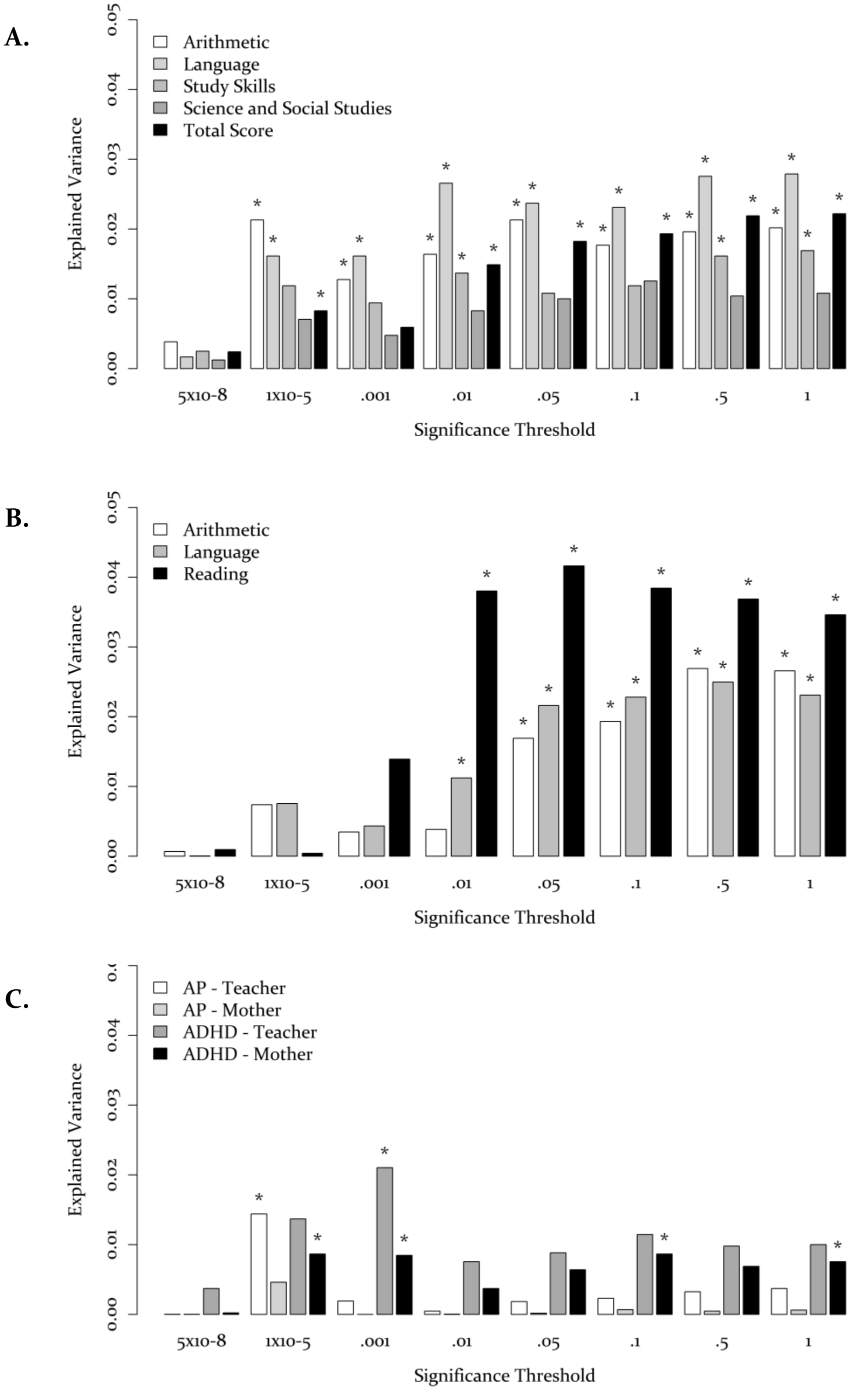


FIGURE 2 The variance explained (R^2) in Educational Achievement, School Performance and Attention Problems (AP)/ADHD symptoms by polygenic scores based on the clumped results for completion of college degree (College) calculated with different p-value thresholds



DISCUSSION

Up to ~3 per cent of the variance in educational achievement and up to ~4 per cent of school performance in children was explained by the polygenic scores that were based on educational attainment in adults. The polygenic scores predicted, in general, educational achievement in children to the same extent as educational attainment in adults (Rietveld et al.,2013). This finding is consistent with numerous results from the genetics literature on general cognitive ability, in which the involvement of the same genetic variants is found in childhood and adulthood (Franic et al.,2014; Haworth et al.,2010; Hoekstra, Bartels and Boomsma,2007). The current study suggests that this is also true for educational achievement, as polygenic scores based on a general measure of educational attainment obtained in adults predicted children's performance across a number of educational domains.

This is the first study that demonstrated genetic associations between ADHD and educational achievement using polygenic scores. The polygenic scores, representing educational attainment in adults, explained up to ~2 per cent of the variance in attention problems and ADHD symptoms in 12-year-olds, indicating that some of the genetic variants that have a positive influence on educational attainment have a protective effect against ADHD. This confirms, at the measured genotype level, the existence of shared genetic effects accounting for the negative association between ADHD and educational achievement, previously found in twin research (Polderman et al.,2010; Saudino and Plomin,2007). This finding implies that at least some of the knowledge obtained with molecular genetic studies towards educational achievement can be used in research towards ADHD.

On the bases of the findings from the polygenic score analyses we cannot establish whether the observed effects of the genetic variants on both educational achievement and ADHD are direct or indirect. It could be that shared underlying biological mechanisms are the cause of the association between educational achievement and ADHD symptoms. When there is pleiotropy some of the genetic variants have a true direct biological influence on both phenotypes. For example, a number of molecular genetic studies demonstrated that there are some genetic variants on chromosome 6, 13 and 14 that have an effect on both reading disability and ADHD (Gayan et al.,2005; Willcutt et al.,2002). The pleiotropic effect of a genetic variant can occur when a gene is involved in multiple biological pathways or the same biological pathway has different effects on the associated phenotypes (Solovieff et al.,2013). For example, dysfunction in the dopaminergic pathway has been implicated in the development of ADHD and this pathway has also been associated with cognitive function (Nieoullon,2002). Alternatively, it may be that the genetic association appears because there are genetic variants influencing ADHD and, being genetically predisposed to ADHD makes it harder to concentrate at school, leading to lower educational achievement. Or, the other way around, children who have problems keeping up in school display, perhaps out of boredom and frustration, ADHD symptoms. The latter hypothesis seems to be refuted by findings from earlier studies that demonstrated that ADHD symptoms measured before children attended primary school are already associated with lower educational achievement years later (Fantuzzo et al.,2003; McGee et al.,1991).

The association between the polygenic scores and ADHD symptoms depended on the measurement instrument and rater. The effects were more strongly related, especially for the mother ratings, to the ADHD index of the CTRS-R/CPRS-R and not to the AP scale of the TRF/CBCL. One possible explanation is that the items of the AP scales are less school-oriented and include items less related to ADHD symptoms as described in the DSM than the ADHD index (Achenbach,1991; American Psychiatric Association,2000; Conners et al.,1998).

The proportion of variance explained by the polygenic scores is small. According to a series of power analyses, this is as expected given the small effects of the individual genetic variants attributed to the complexity of the phenotypes (Dudbridge,2013). It is also consistent with findings in similar studies using polygenic scores to predict other complex phenotypes, for example, schizophrenia (Purcell et al.,2007) and intelligence (Davies et al.,2011). Due to the highly polygenic nature of these phenotypes, the effects of the genetic variants are small and the standard errors of the estimates of the effect sizes are relatively large, limiting the predictive power of the polygenic score. Furthermore, the idea behind polygenic scores is that all genetic variants, also the ones that are not robustly associated with the phenotype, are included in the prediction, resulting in a majority of genetic variants without a true effect. To conclude, the findings of shared genetic variants between educational achievement and ADHD may lead to new possibilities in the research towards ADHD. Larger sample sizes to identify genetic variants influencing ADHD might be within reach, since data on educational achievement are already available at a larger scale for participants with genotype data as well as easier to collect than ADHD data.

APPENDIX

The following people who are not listed as co-authors on this manuscript contributed to the original GWAS meta-analysis on educational attainment (Rietveld et al., 2013), on which the present paper is based. Data access has been granted under section 4 of the Data Sharing Agreement of the Social Science Genetic Association Consortium (SSGAC). The views presented in the present paper may not reflect the opinions of the individuals listed below. The SSGAC is grateful to the authors of (Rietveld et al., 2013) for providing the meta-analysis data. We thank: Abdel Abdellaoui, Arpana Agrawal, Eva Albrecht, Behrooz Z. Alizadeh, Jüri Allik, Najaf Amin, John R. Attia, Stefania Bandinelli, John Barnard, François Bastardot, Sebastian E. Baumeister, Jonathan Beauchamp, Daniel J. Benjamin, Kelly S. Benke, David A. Bennett, Klaus Berger, Lawrence F. Bielak, Laura J. Bierut, Jeffrey A. Boatman, Dorret I. Boomsma, Patricia A. Boyle, Ute Bültmann, Harry Campbell, David Cesarini, Christopher F. Chabris, Lynn Cherkas, Mina K. Chung, Dalton Conley, Francesco Cucca, George Davey-Smith, Gail Davies, Mariza de Andrade, Philip L. De Jager, Christiaan de Leeuw, Jan-Emmanuel De Neve, Ian J. Deary, George V. Dedoussis, Panos Deloukas, Jaime Derringer, Maria Dimitriou, Gudny Eiriksdottir, Niina Eklund, Martin F. Elderson, Johan G. Eriksson, Tõnu Esko, Daniel S. Evans, David M. Evans, Jessica D. Faul, Rudolf Fehrmann, Luigi Ferrucci, Krista Fischer, Lude Franke, Melissa E. Garcia, Christian Gieger, Håkon K. Gjessing, Patrick J.F. Groenen, Henrik Grönberg, Vilmundur Gudnason, Sara Hägg, Per Hall, Jennifer R. Harris, Juliette M. Harris, Tamara B. Harris, Nicholas D. Hastie, Caroline Hayward, Andrew C. Heath, Dena G. Hernandez, Wolgang Hoffmann, Adriaan Hofman, Albert Hofman, Rolf Holle, Elizabeth G. Holliday, Christina Holzapfel, Jouke-Jan Hottenga, William G. Iacono, Carla A. Ibrahim-Verbaas, Thomas Illig, Erik Ingelsson, Bo Jacobsson, Marjo-Riitta Järvelin, Min A. Jhun, Magnus Johannesson, Peter K. Joshi, Astanand Jugessur, Marika Kaakinen, Mika Kähönen, Stavroula Kanoni, Jaakkko Kaprio, Sharon L.R. Kardia, Juha Karjalainen, Robert M. Kirkpatrick, Philipp D. Koellinger, Ivana Kolcic, Matthew Kowgier, Kati Kristiansson, Robert F. Krueger, Zóltan Kutalik, Jari Lahti, David Laibson, Antti Latvala, Lenore J. Launer, Debbie A. Lawlor, Sang H. Lee, Terho Lethimäki, Jingmei Li, Paul Lichtenstein, Peter K. Lichtner, David C. Liewald, Peng Lin, Penelope A. Lind, Yongmei Liu, Kurt Lohman, Marisa Loitfelder, Pamela A. Madden, Patrick K.E. Magnusson, Tomi E. Mäkinen, Pedro Marques Vidal, Nicolas W. Martin, Nicholas G. Martin, Marco Masala, Matt McGue, George McMahon, Sarah E. Medland, Osorio Meirelles, Andres Metspalu, Michelle N. Meyer, Andreas Mielck, Lili Milani, Michael B. Miller, Grant W. Montgomery, Sutapa Mukherjee, Ronny Myhre, Marja-Liisa Nuotio, Dale R. Nyholt, Christopher J. Oldmeadow, Ben A. Oostra, Lyle J. Palmer, Aarno Palotie,

Brenda Penninx, Markus Perola, Katja E. Petrovic, Wouter J. Peyrot, Patricia A. Peyser, Ozren Polašek, Danielle Posthuma, Martin Preisig, Lydia Quaye, Katri Räikkönen, Olli T. Raitakari, Anu Realo, Eva Reinmaa, John P. Rice, Cornelius A. Rietveld, Susan M. Ring, Samuli Ripatti, Fernando Rivadeneira, Thais S. Rizzi, Igor Rudan, Aldo Rustichini, Veikko Salomaa, Antti-Pekka Sarin, David Schlessinger, Helena Schmidt, Reinhold Schmidt, Rodney J. Scott, Konstantin Shakhbazov, Albert V. Smith, Jennifer A. Smith, Harold Snieder, Beate St Pourcain, John M. Starr, Jae Hoon Sul, Ida Surakka, Rauli Svento, Toshiko Tanaka, Antonio Terracciano, Alexander Teumer, A. Roy Thurik, Henning Tiemeier, Nicholas J. Timpson, André G. Uitterlinden, Matthijs J.H.M. van der Loos, Cornelia M. van Duijn, Frank J.A. van Rooij, David R. Van Wagoner, Erkki Vartiainen, Jorma Viikari, Peter M. Visscher, Veronique Vitart, Peter K. Vollenweider, Henry Völzke, Judith M. Vonk, Gérard Waeber, David R. Weir, Jürgen Wellmann, Harm-Jan Westra, H.-Erich Wichmann, Elisabeth Widen, Gonneke Willemsen, James F. Wilson, Alan F. Wright, Jian Yang, Lei Yu, Wei Zhao.

REFERENCES

Abdellaoui, A., Hottenga, J.J., de, K.P., Nivard, M.G., Xiao, X., Scheet, P., Brooks, A., Ehli, E.A., Hu, Y., Davies, G.E., Hudziak, J.J., Sullivan, P.F., van, B.T., Willemsen, G., de Geus, E.J., Penninx, B.W. & Boomsma, D.I. (2013). Population structure, migration, and diversifying selection in the Netherlands. *European Journal of Human Genetics, 21 (11)*, p. 1277-1285.

Achenbach, T.M. (1991). Manual for the Child Behavior Checklist/4 - 18 and 1991 Profile. Burlington, VT: University of Vermont Department of Psychiatry.

American Psychiatric Association (2000). Diagnostic and Statistical Manual of Mental Disorders: 4th ed., text rev. DSM-IV-TR. Washington, DC: APA.

Bartels, M., Rietveld, M.J., Van Baal, G.C. & Boomsma, D.I. (2002). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Research, 5 (6)*, p. 544-553.

Biederman, J., Faraone, S., Milberger, S., Guite, J., Mick, E., Chen, L., Mennin, D., Marrs, A., Ouellette, C., Moore, P., Spencer, T., Norman, D., Wilens, T., Kraus, I. & Perrin, J. (1996). A prospective 4-year follow-up study of attention-deficit hyperactivity and related disorders. *Archives of General Psychiatry, 53 (5)*, p. 437-446.

Boomsma, D.I., de Geus, E.J., Vink, J.M., Stubbe, J.H., Distel, M.A., Hottenga, J.J., Posthuma, D., van Beijsterveldt, T.C., Hudziak, J.J., Bartels, M. & Willemsen, G. (2006). Netherlands Twin Register: from twins to twin families. *Twin Research and Human Genetics, 9 (6)*, p. 849-857.

Cito (2002). Eindtoets Basisonderwijs. Arnhem, The Netherlands: Cito.

Conners, C.K., Sitarenios, G., Parker, J.D. & Epstein, J.N. (1998). Revision and restandardization of the Conners Teacher Rating Scale (CTRS-R): factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology, 26 (4)*, p. 279-291.

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S.E., Liewald, D., Ke, X., Le, H.S., Christoforou, A., Luciano, M., McGhee, K., Lopez, L., Gow, A.J., Corley, J., Redmond, P., Fox, H.C., Haggarty, P., Whalley, L.J., McNeill, G., Goddard, M.E., Espeseth, T., Lundervold, A.J., Reinvang, I., Pickles, A., Steen, V.M., Ollier, W., Porteous, D.J., Horan, M., Starr, J.M., Pendleton, N., Visscher, P.M. & Deary, I.J. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry, 16 (10)*, p. 996-1005.

Deary, I.J., Yang, J., Davies, G., Harris, S.E., Tenesa, A., Liewald, D., Luciano, M., Lopez, L.M., Gow, A.J., Corley, J., Redmond, P., Fox, H.C., Rowe, S.J., Haggarty, P., McNeill, G., Goddard, M.E., Porteous, D.J., Whalley, L.J., Starr, J.M. & Visscher, P.M. (2012). Genetic contributions to stability and change in intelligence from childhood to old age. *Nature, 482 (7384)*, p. 212-215.

Derks, E.M., Hudziak, J.J., Dolan, C.V., van Beijsterveldt, T.C., Verhulst, F.C. & Boomsma, D.I. (2008). Genetic and environmental influences on the relation between attention problems and attention deficit hyperactivity disorder. *Behavior Genetics, 38 (1)*, p. 11-23.

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS.Genet., 9 (3)*, p. e1003348.

Fantuzzo, J., Bulotsky, R., McDermott, P., Mosca, S. & Lutz, M.N. (2003). A multivariate analysis of emotional and behavioral adjustment and preschool educational outcomes. *School Psychology Review, 32 (2),* p. 185-203.

Faraone, S.V., Perlis, R.H., Doyle, A.E., Smoller, J.W., Goralnick, J.J., Holmgren, M.A. & Sklar, P. (2005). Molecular genetics of attention-deficit/hyperactivity disorder. *Biological Psychiatry, 57 (11)*, p. 1313-1323.

Flint, J. & Munafo, M. (2013). Genetics. Herit-ability. *Science, 340 (6139)*, p. 1416-1417.

Franic, S., Dolan, C.V., van Beijsterveldt, C.E., Pol, H.E.H., Bartels, M. & Boomsma, D.I. (2014). Genetic and Environmental Stability of Intelligence in Childhood and Adolescence. *Twin Research and Human Genetics, 17 (3)*, p. 151-163.

Frazier, T.W., Youngstrom, E.A., Glutting, J.J. & Watkins, M.W. (2007). ADHD and achievement: meta-analysis of the child, adolescent, and adult literatures and a concomitant study with college students. *Journal of Learning Disabilities, 40 (1)*, p. 49-65.

Gayan, J., Willcutt, E.G., Fisher, S.E., Francks, C., Cardon, L.R., Olson, R.K., Pennington, B.F., Smith, S.D., Monaco, A.P. & Defries, J.C. (2005). Bivariate linkage scan for reading disability and attention-deficit/hyperactivity disorder localizes pleiotropic loci. *Journal of Child Psychology and Psychiatry, 46 (10)*, p. 1045-1056.

Haworth, C.M., Wright, M.J., Luciano, M., Martin, N.G., de Geus, E.J., van Beijsterveldt, C.E., Bartels, M., Posthuma, D., Boomsma, D.I., Davis, O.S., Kovas, Y., Corley, R.P., Defries, J.C., Hewitt, J.K., Olson, R.K., Rhea, S.A., Wadsworth, S.J., Iacono, W.G., McGue, M., Thompson, L.A., Hart, S.A., Petrill, S.A., Lubinski, D. & Plomin, R. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry, 15 (11)*, p. 1112-1120.

Hoekstra, R.A., Bartels, M. & Boomsma, D.I. (2007). Longitudinal genetic study of verbal and nonverbal IQ from early childhood to young adulthood. *Learning and Individual Differences, 17 (2)*, p. 97-114.

IBM Corp (2011). SPSS Statistics for Windows Version 20.0. Armonk, NY: IBM Corp.

Jöreskog, K. & Sörbom, D. (2002). PRELIS 2 user's reference guide. Chicago Scientific Software.

Julian, T. & Kominski, R. (2011). Education and synthetic work-life earnings estimates. United States: Census Bureau.

Kovas, Y., Haworth, C.M., Dale, P.S. & Plomin, R. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for Research in Child Development, 72 (3)*, p. vii, 1-vii144.

Kuntsi, J., Eley, T.C., Taylor, A., Hughes, C., Asherson, P., Caspi, A. & Moffitt, T.E. (2004). Co-occurrence of ADHD and low IQ has genetic origins. *American Journal of Medical Genetics B. Neuropsychiatric Genetics, 124B (1)*, p. 41-47.

Lencz, T., Knowles, E., Davies, G., Guha, S., Liewald, D.C., Starr, J.M., Djurovic, S., Melle, I., Sundet, K., Christoforou, A., Reinvang, I., Mukherjee, S., Derosse, P., Lundervold, A., Steen, V.M., John, M., Espeseth, T., Raikkonen, K., Widen, E., Palotie, A., Eriksson, J.G., Giegling, I., Konte, B., Ikeda, M., Roussos, P., Giakoumaki, S., Burdick, K.E., Payton, A., Ollier, W., Horan, M., Donohoe, G., Morris, D., Corvin, A., Gill, M., Pendleton, N., Iwata, N., Darvasi, A., Bitsios, P., Rujescu, D., Lahti, J., Hellard, S.L., Keller, M.C., Andreassen, O.A., Deary, I.J., Glahn, D.C. and Malhotra, A.K. (2014). Molecular genetic evidence for overlap between general cognitive ability and risk for schizophrenia: a report from the Cognitive Genomics consorTium (COGENT). *Molecular Psychiatry, 19 (2)*, p. 168-174.

Li, J. & Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity, 95 (3)*, p. 221-227.

Mackenbach, J.P., Kunst, A.E., Cavelaars, A.E., Groenhof, F. & Geurts, J.J. (1997). Socioeconomic inequalities in morbidity and mortality in western Europe. The EU Working Group on Socioeconomic Inequalities in Health. *Lancet, 349 (9066)*, p. 1655-1659.

McGee, R., Partridge, F., Williams, S. & Silva, P.A. (1991). A twelve-year follow-up of preschool hyperactive children. *Journal of the American Academy of Child and Adolescent Psychiatry, 30 (2)*, p. 224-232.

Mick, E., Todorov, A., Smalley, S., Hu, X., Loo, S., Todd, R.D., Biederman, J., Byrne, D., Dechairo, B., Guiney, A., McCracken, J., McGough, J., Nelson, S.F., Reiersen, A.M., Wilens, T.E., Wozniak, J., Neale, B.M. & Faraone, S.V. (2010). Family-based genome-wide association scan of attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 49 (9)*, p. 898-905.

Neale, B.M., Lasky-Su, J., Anney, R., Franke, B., Zhou, K., Maller, J.B., Vasquez, A.A., Asherson, P., Chen, W., Banaschewski, T., Buitelaar, J., Ebstein, R., Gill, M., Miranda, A., Oades, R.D., Roeyers, H., Rothenberger, A., Sergeant, J., Steinhausen, H.C., Sonuga-Barke, E., Mulas, F., Taylor, E., Laird, N., Lange, C., Daly, M. & Faraone, S.V. (2008). Genome-wide association scan of attention deficit hyperactivity disorder. *American Journal of Medical Genetics B. Neuropsychiatric Genetics, 147B (8)*, p. 1337-1344.

Neale, B.M., Medland, S., Ripke, S., Anney, R.J., Asherson, P., Buitelaar, J., Franke, B., Gill, M., Kent, L., Holmans, P., Middleton, F., Thapar, A., Lesch, K.P., Faraone, S.V., Daly, M., Nguyen, T.T., Schafer, H., Steinhausen, H.C., Reif, A., Renner, T.J., Romanos, M., Romanos, J., Warnke, A., Walitza, S., Freitag, C., Meyer, J., Palmason, H., Rothenberger, A., Hawi, Z., Sergeant, J., Roeyers, H., Mick, E. & Biederman, J. (2010a). Case-control genome-wide association study of attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 49 (9)*, p. 906-920.

Neale, B.M., Medland, S.E., Ripke, S., Asherson, P., Franke, B., Lesch, K.P., Faraone, S.V., Nguyen, T.T., Schafer, H., Holmans, P., Daly, M., Steinhausen, H.C., Freitag, C., Reif, A., Renner, T.J., Romanos, M., Romanos, J., Walitza, S., Warnke, A., Meyer, J., Palmason, H., Buitelaar, J., Vasquez, A.A., Lambregts-Rommelse, N., Gill, M., Anney, R.J., Langely, K., O'Donovan, M., Williams, N., Owen, M., Thapar, A., Kent, L., Sergeant, J., Roeyers, H., Mick, E., Biederman, J., Doyle, A., Smalley, S., Loo, S., Hakonarson, H., Elia, J., Todorov, A., Miranda, A., Mulas, F., Ebstein, R.P., Rothenberger, A., Banaschewski, T., Oades, R.D., Sonuga-Barke, E., McGough, J., Nisenbaum, L., Middleton, F., Hu, X. & Nelson, S. (2010b). Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 49 (9)*, p. 884-897.

Nieoullon, A. (2002). Dopamine and the regulation of cognition and attention. *Progress in Neurobiology, 67 (1)*, p. 53-83.

Nyholt, D.R. (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *American Journal of Human Genetics, 74 (4)*, p. 765-769.

Polderman, T.J., Gosso, M.F., Posthuma, D., van Beijsterveldt, T.C., Heutink, P., Verhulst, F.C. & Boomsma, D.I. (2006). A longitudinal twin study on IQ, executive functioning, and attention problems during childhood and early adolescence. *Acta Neurologica Belgica, 106 (4)*, p. 191-207.

Polderman, T.J.C., Boomsma, D.I., Bartels, M., Verhulst, F.C. & Huizink, A.C. (2010). A systematic review of prospective studies on attention problems and academic achievement. *Acta Psychiatrica Scandinavica, 122 (4)*, p. 271-284.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. & Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics, 81 (3)*, p. 559-575.

Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F. & Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature, 460 (7256)*, p. 748-752.

Rietveld, C.A., Medland, S.E., Derringer, J., Yang, J., Esko, T., Martin, N.W., Westra, H.J., Shakhbazov, K., Abdellaoui, A., Agrawal, A., Albrecht, E., Alizadeh, B.Z., Amin, N., Barnard, J., Baumeister, S.E., Benke, K.S., Bielak, L.F., Boatman, J.A., Boyle, P.A., Davies, G., de, L.C., Eklund, N., Evans, D.S., Ferhmann, R., Fischer, K., Gieger, C., Gjessing, H.K., Hagg, S., Harris, J.R., Hayward, C., Holzapfel, C., Ibrahim-Verbaas, C.A., Ingelsson, E., Jacobsson, B., Joshi, P.K., Jugessur, A., Kaakinen, M., Kanoni, S., Karjalainen, J., Kolcic, I., Kristiansson, K., Kutalik, Z., Lahti, J., Lee, S.H., Lin, P., Lind, P.A., Liu, Y., Lohman, K., Loitfelder, M., McMahon, G., Vidal, P.M., Meirelles, O., Milani, L., Myhre, R., Nuotio, M.L., Oldmeadow, C.J., Petrovic, K.E., Peyrot, W.J., Polasek, O., Quaye, L., Reinmaa, E., Rice, J.P., Rizzi, T.S., Schmidt, H., Schmidt, R., Smith, A.V., Smith, J.A., Tanaka, T., Terracciano, A., van der Loos, M.J., Vitart, V., Volzke, H., Wellmann, J., Yu, L., Zhao, W., Allik, J., Attia, J.R., Bandinelli, S., Bastardot, F., Beauchamp, J., Bennett, D.A., Berger, K., Bierut, L.J., Boomsma, D.I., Bultmann, U., Campbell, H., Chabris, C.F., Cherkas, L., Chung, M.K., Cucca, F., de, A.M., De Jager, P.L., De Neve, J.E., Deary, I.J., Dedoussis, G.V., Deloukas, P., Dimitriou, M., Eiriksdottir, G., Elderson, M.F., Eriksson, J.G., Evans, D.M., Faul, J.D., Ferrucci, L., Garcia, M.E., Gronberg, H., Guethnason, V., Hall, P., Harris, J.M., Harris, T.B., Hastie, N.D., Heath, A.C., Hernandez, D.G., Hoffmann, W., Hofman, A., Holle, R., Holliday, E.G., Hottenga, J.J., Iacono, W.G., Illig, T., Jarvelin, M.R., Kahonen, M., Kaprio, J., Kirkpatrick, R.M., Kowgier, M., Latvala, A., Launer, L.J., Lawlor, D.A., Lehtimaki, T., Li, J., Lichtenstein, P., Lichtner, P., Liewald, D.C., Madden, P.A., Magnusson, P.K., Makinen, T.E., Masala, M., McGue, M., Metspalu, A., Mielck, A., Miller, M.B., Montgomery, G.W., Mukherjee, S., Nyholt, D.R., Oostra, B.A., Palmer, L.J., Palotie, A., Penninx, B.W., Perola, M., Peyser, P.A., Preisig, M., Raikkonen, K., Raitakari, O.T., Realo, A., Ring, S.M., Ripatti, S., Rivadeneira, F., Rudan, I., Rustichini, A., Salomaa, V., Sarin, A.P., Schlessinger, D., Scott, R.J., Snieder, H., St, P.B., Starr, J.M., Sul, J.H., Surakka, I., Svento, R., Teumer, A., Tiemeier, H., van Rooij, F.J., Van Wagoner, D.R., Vartiainen, E., Viikari, J., Vollenweider, P., Vonk, J.M., Waeber, G., Weir, D.R., Wichmann, H.E., Widen, E., Willemsen, G., Wilson, J.F., Wright, A.F., Conley, D., Davey-Smith, G., Franke, L., Groenen, P.J., Hofman, A., Johannesson, M., Kardia, S.L., Krueger, R.F., Laibson, D., Martin, N.G., Meyer, M.N., Posthuma, D., Thurik, A.R., Timpson, N.J., Uitterlinden, A.G., van Duijn, C.M., Visscher, P.M., Benjamin, D.J., Cesarini, D. & Koellinger, P.D. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science, 340 (6139)*, p. 1467-1471.

Saudino, K.J. & Plomin, R. (2007). Why are hyperactivity and academic achievement related? *Child Development, 78 (3)*, p. 972-986.

Scheet, P., Ehli, E.A., Xiao, X., van Beijsterveldt, C.E., Abdellaoui, A., Althoff, R.R., Hottenga, J.J., Willemsen, G., Nelson, K.A., Huizenga, P.E., Hu, Y., Amos, C.I., Bartels, M., Groen-Blokhuis, M.M., de Geus, E.J., Hudziak, J.J., Davies, G.E. & Boomsma, D.I. (2012). Twins, tissue, and time: an assessment of SNPs and CNVs. *Twin Research and Human Genetics, 15 (6)*, p. 737-745.

Solovieff, N., Cotsapas, C., Lee, P.H., Purcell, S.M. & Smoller, J.W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics, 14 (7)*, p. 483-495.

Stergiakouli, E., Hamshere, M., Holmans, P., Langley, K., Zaharieva, I., Hawi, Z., Kent, L., Gill, M., Williams, N., Owen, M.J., O'Donovan, M. & Thapar, A. (2012). Investigating the contribution of common genetic variants to the risk and pathogenesis of ADHD. *American Journal of Psychiatry, 169 (2)*, p. 186-194.

van Beijsterveldt, C.E., Groen-Blokhuis, M., Hottenga, J.J., Franic, S., Hudziak, J.J., Lamb, D., Huppertz, C., de Zeeuw, E., Nivard, M., Schutte, N., Swagerman, S., Glasner, T., van, F.M., Brouwer, C., Stroet, T., Nowotny, D., Ehli, E.A., Davies, G.E., Scheet, P., Orlebeke, J.F., Kan, K.J., Smit, D., Dolan, C.V., Middeldorp, C.M., de Geus, E.J., Bartels, M. & Boomsma, D.I. (2013). The Young Netherlands Twin Register (YNTR): longitudinal twin and family studies in over 70,000 children. *Twin Research and Human Genetics, 16 (1)*, p. 252-267.

van Dijk, B.A., Boomsma, D.I. & de Man, A.J. (1996). Blood group chimerism in human multiple births is not rare. *American Journal of Medical Genetics, 61 (3)*, p. 264-268.

van Soelen, I.L., Brouwer, R.M., Peper, J.S., van, L.M., Koenis, M.M., van Beijsterveldt, T.C., Swagerman, S.C., Kahn, R.S., Hulshoff Pol, H.E. & Boomsma, D.I. (2012). Brain SCALE: brain structure and cognition: an adolescent longitudinal twin study into the genetic etiology of individual differences. *Twin Research and Human Genetics, 15 (3)*, p. 453-467.

Willcutt, E.G., Pennington, B.F., Smith, S.D., Cardon, L.R., Gayan, J., Knopik, V.S., Olson, R.K. & Defries, J.C. (2002). Quantitative trait locus for reading disability on chromosome 6p is pleiotropic for attention-deficit/hyperactivity disorder. *American Journal of Medical Genetics, 114 (3)*, p. 260-268.

Willer, C.J., Li, Y. & Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics, 26 (17)*, p. 2190-2191.

8.

LONGITUDINAL MODELING SUGGESTS THAT ADHD SYMPTOMS LEAD TO LOWER EDUCATIONAL ACHIEVEMENT IN CHILDREN

Based on Eveline L. de Zeeuw, Catharina E.M. van Beijsterveldt, Eco J. C. de Geus and Dorret I. Boomsma. (2015). Longitudinal genetic causal modeling suggests that ADHD symptoms lead to lower educational achievement in children. In Preparation.

During childhood there is a negative correlation between Oppositional Defiant Disorder (ODD) and educational achievement and between Attention Deficit Hyperactivity Disorder (ADHD) and educational achievement. Twin studies suggest that the genetic factors influencing ODD/ADHD and educational achievement are also correlated. A genetic correlation can reflect pleiotropy, or can be induced by a causal effect of ODD and ADHD symptoms on educational achievement. In this study, the hypothesis of a causal effect is tested against the hypothesis of genetic pleiotropy using a genetically sensitive design. Complete data on ODD and ADHD symptoms and on educational achievement were available in a cross-sectional sample of 8789 children and in a longitudinal sample of 4540 children, registered with the Netherlands Twin Register. In both sexes, more ODD (boys: r = -.08; girls: r = -.09) and ADHD (boys: r = -.16 to -.39; girls: r = -.12 to -.41) symptoms were associated with lower educational achievement. The observed longitudinal associations were of a similar magnitude. Comparing differences between children from genetically identical twin pairs, girls with more ODD symptoms had lower educational achievement than their co-twin, but this difference was not seen in boys. The twin with more ADHD symptoms scored significantly lower on educational achievement than the co-twin. All genetic correlations between ODD symptoms and educational achievement were significantly different from zero (boys: r = -.09 to -.11; girls: r = -.09 to -.17) while most environmental correlations were not. In contrast, for ADHD symptoms and educational achievement, all genetic correlations (boys: r = -.20 to -.48; girls: r = -.13 to -.48) and most environmental correlations (boys: r = -.05 to -.33; girls: r = -.05 to -.37) were significant. ADHD symptoms may causally lower educational achievement whereas genetic pleiotropic effects are the most likely cause for the association between ODD symptoms low educational achievement.

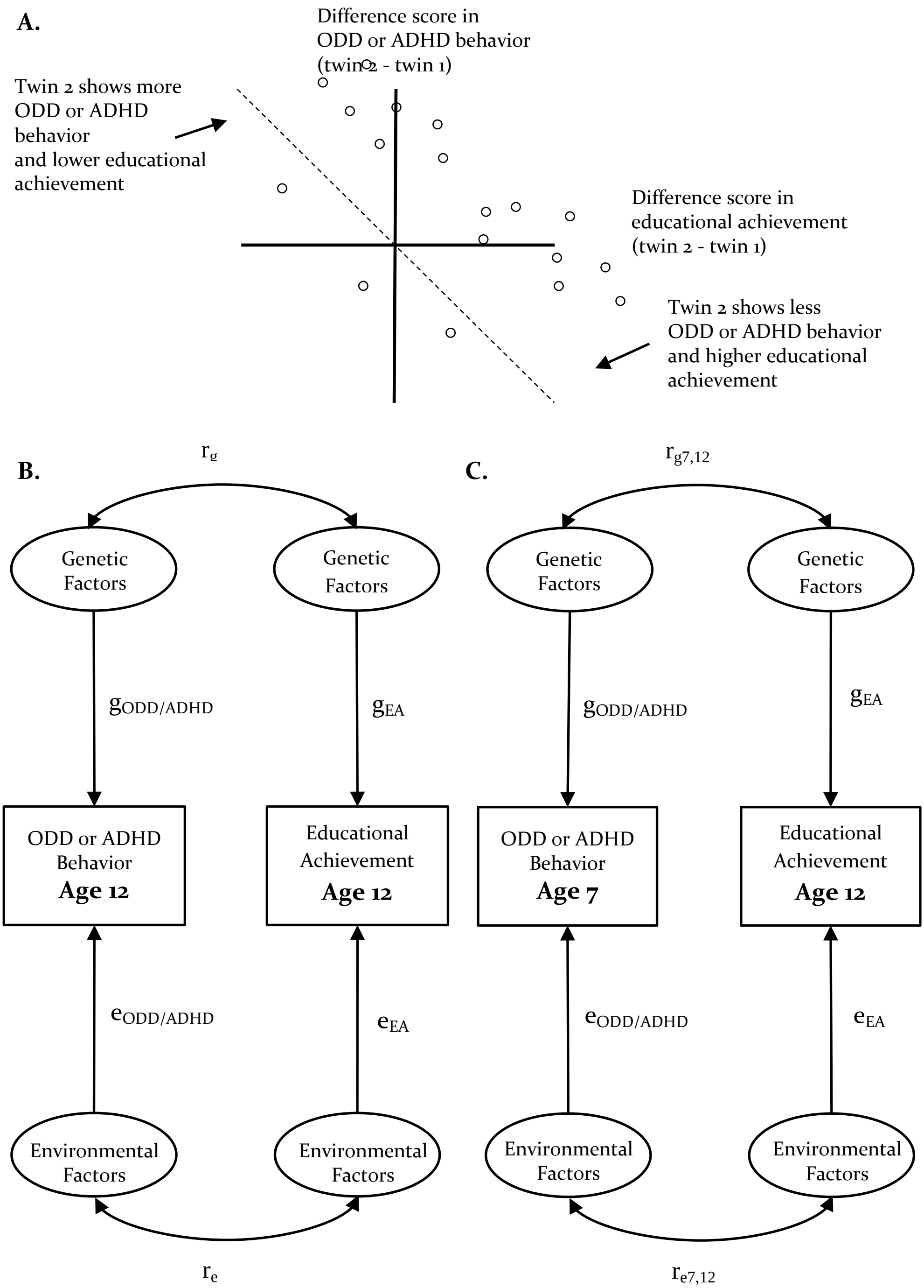
INTRODUCTION

Low educational achievement in children is an important predictor of continued low achievement and school dropout (Moilanen, Shaw & Maxwell, 2010). The American Psychiatric Association (APA) estimates that 3 to 7 per cent of all school aged children are diagnosed with ADHD and that the prevalence of ODD in children is between 2 to 16 per cent (American Psychiatric Association, 2000). More than 50 per cent of the children diagnosed with ADHD also have ODD (Angold, Costello & Erkanli, 1999). It is well recognized that ODD and ADHD diagnoses can be considered the extreme end of the normal distribution of symptoms in the population (Hudziak et al., 2005; Lubke et al., 2009). In both clinical and population samples, there is a significant negative association between ADHD symptoms and educational achievement (Polderman, Boomsma, Bartels, Verhulst & Huizink, 2010). Few studies have looked at the association between ODD and educational achievement.

Behavioral genetic studies have established that ADHD is amongst the most heritable psychiatric childhood disorders. According to a review of 20 twin studies, the mean estimate of the heritability of ADHD in children is over 75 per cent (Faraone et al., 2005). Estimates for ODD are somewhat lower with a heritability of around 50 per cent (Hudziak et al., 2005). Educational achievement shows a relatively consistent etiology with a moderate to high influence of genes and a small influence of common environmental factors (Bartels et al., 2002; Haworth et al., 2011). Twin studies have already shown that the (longitudinal) association between ADHD symptoms and educational achievement can to a large extent be attributed to an overlap in genetic factors (Greven et al., 2014; Kuntsi et al., 2004; Saudino & Plomin, 2007). Recently, the genetic correlation between ADHD symptoms and educational achievement has also been demonstrated by a significant prediction of ADHD symptoms in children by polygenic scores which were based on the effect sizes of genetic variants, measured at the genotype level, from a genome-wide association study towards educational attainment in adults (de Zeeuw et al., 2014). The other way around, polygenic scores on genetic variants associated with clinical ADHD predicted general cognitive ability in the general population (Martin et al., 2014). There are two explanations for the observed genetic correlation between ODD and ADHD symptoms and educational achievement. One explanation is genetic pleiotropy, which is when the same genes through the same underlying biological mechanisms, for example brain processes, affect ODD or ADHD symptoms as well as educational achievement. Another possible explanation for the observed genetic correlation is that there is a causal effect of ODD or ADHD on educational achievement which makes it harder to concentrate at school, eventually leading to lower educational achievement. The genetic variants influencing ODD or ADHD would then, through the causal chain, also influence educational achievement.

In the present study we test, in a large population sample, the hypothesis of a causal effect of ODD and ADHD symptoms on educational achievement against the alternative hypothesis of genetic pleiotropy. There are several testable predictions that follow from the causal hypothesis (de Moor et al, 2008). A first prediction is that within pair differences in ODD and ADHD symptoms should be associated with within pair differences in educational achievement in genetically identical or monozygotic (MZ) twins as it excludes confounding by genes and shared environment by the twins such as childhood neighborhood and parental SES. A child, from a MZ twin pair, who shows more ODD or ADHD symptoms than his or her co-twin should also perform worse in school. A non-significant correlation would point to genetic factors driving the association between ODD or ADHD symptoms and educational achievement while a significant correlation would support the causal hypothesis (Figure 1A). A second prediction is that if ODD or ADHD symptoms have a causal effect on educational achievement, all factors with an effect on ODD or ADHD symptoms should also have an effect on educational achievement. This also holds for the latent genetic and environmental factors detected in a twin study. The correlation between the latent genetic and environmental factors influencing the two phenotypes can be tested in a cross-sectional correlational twin model (Neale, Røysamb & Jacobson, 2006) (Figure 1B). A third, related, prediction is that the association between ODD or ADHD symptoms and educational achievement also exists longitudinally and that the same genetic and environmental factors influencing ODD or ADHD at baseline also influence educational attainment at follow-up. This can be tested in a longitudinal correlational twin model (Neale, Røysamb & Jacobson, 2006) by computing the genetic and environmental correlations over time (Figure 1C). The finding that the genetic correlations as well as the environmental correlations are significant in the cross-sectional and longitudinal models would be in support of, not evidence for, the causal hypothesis. If only the genetic correlations are significant, the causal hypothesis would be rejected, and genetic pleiotropy would be more likely.

FIGURE 1 Graphic representation of the three models, MZ within twin pair differences model (A), cross-sectional correlational model (B) and longitudinal correlational model (C), used to test the causal effect of ODD and ADHD symptoms on educational achievement



METHODS

PARTICIPANTS

The Netherlands Twin Register (NTR), established around 1987 by the department of Biological Psychology at the VU University Amsterdam, registers approximately 40 per cent of all multiple births in the Netherlands. The parents of the twins receive a survey about the development of their children every two years until the twins are 12 years old (Bartels et al., 2007; van Beijsterveldt et al., 2013). The survey sent to the parents includes, amongst others, the short version of the Conners' Parent Ratings Scale - Revised (CPRS-R). In addition, when the children are approximately 12 years old, parents are asked to report the scores of their children on a national test of educational achievement, which is administered in the final grade of primary school (Cito, 2002).

Data on ODD and/or ADHD symptoms are available for age 7 and age 12 while data on educational achievement are only available for age 12. 8789 children had data on ODD and/or ADHD symptoms at age 12 as well as data on educational achievement at age 12. 4406 children only had data on ODD and/or ADHD symptoms at age 12 and 1555 children only had data on educational achievement at age 12 (cross-sectional sample). 4540 children had data on ODD and/or ADHD symptoms at age 7 as well as data on educational achievement at age 12. 8594 children only had data on ODD and/or ADHD symptoms at age 7 and 5804 children only had data on educational achievement at age 12 (longitudinal sample). Children with a disease or handicap that interfered severely with daily functioning were excluded for this study. The cross-sectional sample included 2479 twin pairs of opposite sex. For the same-sex twin pairs, determination of zygosity status was based on blood or DNA polymorphisms (N=1124) or on the basis of parental report of items on resemblance in appearance and confusion of the twins by parents and others (N=3950). The longitudinal sample included 3169 twin pairs of opposite sex. For the same-sex twin pairs, determination of zygosity status was based on blood or DNA polymorphisms (N=1072) or on the basis of the earlier described parental report (N=5510). The parental report establishes zygosity with an accuracy of approximately 93 per cent (Rietveld et al., 2000).

MEASUREMENTS

A national standardized test of educational achievement is administered in the final grade of primary school at approximately 80 per cent of all schools in the Netherlands This test measures what a child has learned across all grades of primary school and is used to give a recommendation about the level of secondary education suitable for the child. The test consists of multiple choice items in four different domains, namely Arithmetic, Language, Study Skills and

Science and Social Studies. The first three test scales are combined into a Total Score, standardized on a scale from 500 and 550, which is used in this study to measure educational achievement.

ODD and ADHD symptoms were assessed, by mothers, with the short version of the Conners' Parent Rating Scales - Revised (CPRS-R). The CPRS-R consists of 27 items scored on a 4 point scale from 0 (not true or never) to 3 (completely true or very often). The CTRS-P includes 4 scales measuring Oppositional Behavior (OPP: 6 items), Inattention (ATT: 6 items), Hyperactivity (HYP: 6 items) and Attention Deficit Hyperactivity Disorder Index (ADHD: 12 items). Three items are included in both the ATT and ADHD scale ('Avoids, or has difficulties in engaging in tasks for a longer period of time', 'Has trouble concentrating in class' and 'Does not follow instructions or finish homework'). Sum scores for the number of symptoms were computed when subjects had no or a limited number of missing items on a scale. Missing items were imputed by the rounded averaged item score of the scale for that child. Sum scores showed an L-shaped distribution and therefore the data were square root transformed prior to analyses.

STATISTICAL ANALYSES

The causal effect of ODD and ADHD symptoms on educational achievement was tested in three different models (de Moor et al, 2008).. The first two testable hypotheses were based on cross-sectional data, the MZ twin within pair differences model and the cross-sectional correlational model, while the third one, the longitudinal correlational model, was based on longitudinal data.

For the first hypothesis, the difference in ODD or ADHD symptoms and the difference in educational achievement between twins from MZ twin pairs were computed and correlated within the twin pairs in SPSS 21.0 (IBM Corp, 2012). The use of data from MZ twin pairs removes possible confounding by genes and shared environment since the MZ twin pairs are genetically identical and grow up in partly the same environment. A significant correlation between ODD or ADHD symptoms and educational achievement would indicate that the association is not merely due to genes or shared environment and would support a causal hypothesis, whereas a non-significant correlation would support genetic pleiotropy.

For the second and third hypothesis, it was assessed whether the cross-sectional association between ODD or ADHD symptoms at age 12 and educational achievement at age 12 and the longitudinal association between ODD or ADHD at age 7 and educational achievement at age 12 were paralleled by significant genetic and environmental correlations.

Cross-sectional and longitudinal models were fitted to the data in the R (R Core Team, 2014) package OpenMx Version 3.1.0 (Boker S.M. et al., 2011; Boker et al., 2012) using raw data maximum likelihood estimation. The analyses were run separately for each scale of the CPRS-R. A model that freely estimated all parameters, i.e. means, variances and covariances, separately for the different zygosity-by-gender groups (MZm, DZm, MZf, DZf and DOS), was fitted to the data (saturated model).

The difference in resemblance between monozygotic (MZ), sharing (nearly) all genes, and dizygotic (DZ) twin pairs, sharing approximately 50 per cent of their segregating genes, was used to estimate the contribution of genes (heritability) and the environment to the different phenotypes. Similarly to the decomposition of variance in a univariate model (Plomin et al., 2008), the cross-twin and cross-phenotype correlation between MZ and DZ twin pairs forms the basis to estimate the genetic and the environmental correlations between phenotypes.

Genetic and environmental correlations were estimated in a series of bivariate genetic models, which included three latent factors, i.e. additive genetic factors (A), common environmental (C) and unique (E) environmental factors (Neale, Røysamb & Jacobson, 2006). Estimates for the influence of the latent factors on ODD symptoms, ADHD symptoms and educational achievement, were estimated separately for boys and girls. Means were allowed to be different between boys and girls for all phenotypes. A causal effect of ODD and ADHD symptoms on educational achievement implies that all genetic and environmental factors influencing ODD and ADHD symptoms affect educational achievement. This implies that both genetic and environmental correlations should be significant.

Significance testing was done by constraining parameter values at zero and comparing the fit of the submodel to that of the unconstrained model. Testing the significance of the correlation between the common environmental factors is only possible when individual differences in both ODD or ADHD symptoms and educational achievement are influenced by the common environment. If there is no variation in ODD or ADHD symptoms accounted for by common environmental factors there will be no common environmental correlation between ODD or ADHD symptoms and educational achievement. In this case, the causal hypothesis will be supported when both the genetic and unique environmental correlation are significant. The difference in goodness of fit between nested models was assessed by log-likelihood ratio tests (LRT) which calculate the difference in -2log-likelihood (-2LL) between two models and evaluates this χ^2-statistic with the difference in the number of estimated parameters between the models as degrees of freedom. A p-value smaller than 0.05 was considered significant.

RESULTS

Table 1 presents the means and standard deviations for boys and girls, for ODD and ADHD symptoms at age 7 and 12 and educational achievement at age 12. The cross-sectional and longitudinal phenotypic correlations between ODD or ADHD symptoms and educational achievement are given in Table 2. These correlations tend to be small for ODD symptoms, but all estimates were significant. The cross-sectional and longitudinal phenotypic correlations between ADHD symptoms and educational achievement are larger, and also more variable, depending on the subtype of symptoms. The longitudinal phenotypic correlations are rather similar to the cross-sectional phenotypic correlations.

TABLE 1 Means and standard deviations (SD) for ODD and ADHD symptoms and educational achievement

	Boys			Girls		
	N	**Mean**	**SD**	**N**	**Mean**	**SD**
Oppositional Behavior						
Age 7	6559	4.27	3.51	6530	3.61	3.16
Age 12	6487	3.75	3.31	6685	3.30	3.03
Inattention						
Age 7	6524	3.66	4.11	6484	2.44	3.31
Age 12	6484	3.75	4.09	6682	2.32	3.08
Hyperactivity						
Age 7	6562	3.41	3.63	6533	2.03	2.65
Age 12	6488	2.01	2.80	6687	1.06	1.89
ADHD Index						
Age 7	6555	8.47	7.65	6521	5.90	6.17
Age 12	6489	7.67	7.25	6681	4.96	5.59
Educational Achievement						
Age 12	4950	538.5	8.3	5394	537.1	8.7

The correlations for within MZ pair differences in ODD or ADHD symptoms and educational achievement are also reported in Table 2. The correlation was significant for ODD symptoms and educational achievement in girls but not in boys. Girls from MZ pairs with more ODD symptoms perform less in school than their sisters. For boys this is not seen. The correlations between the MZ twin pair differences for ADHD symptoms and educational achievement were significant for both boys and girls. Thus, in genetically identical twin pairs, the twin with more ADHD symptoms has a lower educational achievement than his or her co-twin. As within these twin pairs there is no confounding by genes, the within-pair association between ODD symptoms, and even more so, ADHD symptoms and educational achievement cannot reflect genetic pleiotropy. They therefore more likely reflect causality.

TABLE 2 Monozygotic within twin pair differences correlations, cross-sectional correlations and longitudinal correlations (N) for ODD and ADHD symptoms with educational achievement

	Oppositional Behavior		**Inattention**		**Hyperactivity**		**ADHD Index**	
Cross-sectional Correlation								
Boys	-.08*	(4218)	-.39**	(4214)	-.16**	(4216)	-.32**	(4218)
Girls	-.09*	(4560)	-.41**	(4558)	-.12**	(4561)	-.33**	(4566)
Longitudinal Correlation								
Boys	-.09**	(2149)	-.32**	(2143)	-.16**	(2155)	-.27**	(2146)
Girls	-.11**	(2376)	-.33**	(2359)	-.17**	(2374)	-.28**	(2368)
MZ Differences Correlation								
Boys	-.04	(678)	-.29**	(686)	-.10*	(687)	-.31**	(687)
Girls	-.11*	(834)	-.34**	(835)	-.11*	(838)	-.31**	(836)

* $p < .01$; ** $p < .001$

The twin correlations as estimated in the cross-sectional and longitudinal correlational models are summarized in Table 3 and 4. All MZ within phenotype correlations were larger than the DZ correlations. Most MZ cross-correlations between ODD symptoms and educational achievement and between ADHD symptoms and educational achievement were significant and higher than the DZ cross-correlations. This pattern of correlations suggests that there is a genetic correlation which is at least partly responsible for the cross-sectional and longitudinal associations between ODD and ADHD symptoms and educational achievement.

TABLE 3 Cross-sectional cross-twin within-phenotype (upper) and cross-twin cross-phenotype (lower) correlations (95% Confidence Interval) for ODD and ADHD symptoms at age 12 and educational achievement at age 12

	Educational Achievement	Oppositional Behavior	Inattention	Hyperactivity	ADHD Index
MZm	.81 (.78; .83)	.73 (.70; .76)	.74 (.72; .77)	.82 (.80; .84)	.77 (.74; .79)
DZm	.46 (.39; .51)	.43 (.38; .48)	.29 (.24; .35)	.39 (.34; .43)	.34 (.29; .39)
MZf	.83 (.81; .85)	.70 (.67; .73)	.74 (.71; .76)	.80 (.78; .82)	.76 (.73; .78)
DZf	.43 (.37; .48)	.46 (.41; .50)	.27 (.21; .33)	.42 (.37; .47)	.29 (.23; .34)
DOS	.44 (.40; .48)	.41 (.37; .45)	.27 (.23; .31)	.38 (.35; .42)	.29 (.25; .33)
MZm		-.16 (-.24; -.07)	-.31 (-.39; -.23)	-.17 (-.26; -.08)	-.27 (-.35; -.18)
DZm		-.06 (-.14; .02)	-.06 (-.13; .02)	-.04 (-.12; .04)	-.08 (-.15; .00)
MZf		-.11 (-.19; -.02)	-.22 (-.30; -.13)	-.05 (-.14; .03)	-.19 (-.27; -.10)
DZf		-.05 (-.13; .03)	-.06 (-.13; .02)	-.09 (-.17; -.01)	-.04 (-.12; .04)
DOS		-.07 (-.13; -.02)	-.13 (-.18; -.08)	-.14 (-.20; -.09)	-.12 (-.17; -.06)

MZm = monozygotic boys; DZm = dizygotic boys; MZf = monozygotic girls; DZf = dizygotic girls; DOS = dizygotic of opposite-sex

Table 5 includes the estimates and 95% confidence intervals of the cross-sectional genetic and environmental correlations between ODD or ADHD symptoms at age 12 and educational achievement at age 12. The genetic correlations between ODD symptoms and educational achievement were significant as was the environmental correlation in girls, but the environmental correlation in boys was not significantly different from zero. The genetic correlations and environmental correlations between ADHD symptoms and educational achievement were all significant. This applied to both the ATT and HYP subscales.

TABLE 4 Longitudinal cross-twin within-phenotype (upper) and cross-twin cross-phenotype (lower) correlations (95% Confidence Interval) for ODD and ADHD symptoms at age 7 and educational achievement at age 12

	Educational Achievement	Oppositional Behavior	Inattention	Hyperactivity	ADHD Index
MZm	.81 (.78; .83)	.72 (.69; .74)	.77 (.75; .80)	.78 (.75; .80)	.82 (.80; .84)
DZm	.46 (.39; .51)	.41 (.35; .46)	.23 (.18; .29)	.29 (.23; .34)	.28 (.23; .34)
MZf	.83 (.81; .85)	.74 (.71; .76)	.70 (.67; .72)	.75 (.73; .78)	.75 (.72; .77)
DZf	.43 (.37; .48)	.44 (.38; .48)	.24 (.18; .30)	.35 (.29; .40)	.33 (.27; .38)
DOS	.44 (.40; .48)	.41 (.38; .45)	.19 (.15; .23)	.29 (.25; .33)	.29 (.25; .32)
MZm		-.10 (-.16; -.03)	-.37 (-.42; -.31)	-.14 (-.21; -.08)	-.29 (-.35; -.23)
DZm		-.05 (-.11; .00)	-.07 (-.12; -.01)	-.06 (-.11; -.00)	-.06 (-.12; -.00)
MZf		-.07 (-.13; -.01)	-.32 (-.37; -.27)	-.07 (-.13; -.00)	-.26 (-.31; -.20)
DZf		-.06 (-.12; .00)	-.03 (-.09; .03)	-.02 (-.08; .04)	-.03 (-.09; .03)
DOS		-.06 (-.10; -.02)	-.11 (-.15; -.08)	-.09 (-.13; -.05)	-.09 (-.13; -.06)

MZm = monozygotic boys; DZm = dizygotic boys; MZf = monozygotic girls; DZf = dizygotic girls; DOS = dizygotic of opposite-sex

TABLE 5 Cross-sectional genetic and environmental correlations (95% Confidence Interval), separately for boys and girls, for ODD and ADHD symptoms at age 12 with educational achievement at age 12

	Oppositional Behavior	Inattention	Hyperactivity	ADHD Index
Genetic Correlation				
Boys	-.09 (-.15; -.04)	-.48 (-.54; -.42)	-.20 (-.25; -.15)	-.37 (-.42; -.32)
Girls	-.09 (-.14; -.04)	-.48 (-.53; -.43)	-.13 (-.18; -.08)	-.37 (-.42; -.33)
Environmental Correlation				
Boys	-.02 (-.10; .06)	-.30 (-.37; -.23)	-.10 (-.17; -.02)	-.33 (-.39; -.26)
Girls	-.09 (-.16; -.02)	-.37 (-.43; -.31)	-.11 (-.18; -.04)	-.34 (-.40; -.28)

Table 6 gives the estimates and 95% confidence intervals of the longitudinal genetic and environmental correlations between ODD and ADHD symptoms at age 7 and educational achievement at age 12. The genetic correlations were significant for ODD symptoms but the environmental correlations were not.

The genetic correlations between the different subtypes of ADHD symptoms and educational achievement were all significant whereas the environmental correlations were significant for ATT and ADHD but not HYP. Taken together, table 5 and 6 show that all latent factors influencing ADHD symptoms also influenced current and future educational achievement, in keeping with the predictions from the causal hypothesis.

TABLE 6 Longitudinal genetic and environmental correlations (95% Confidence Interval), separately for boys and girls, for ODD and ADHD symptoms at age 7 with educational achievement at age 12

	Oppositional Behavior	Inattention	Hyperactivity	ADHD Index
Genetic Correlation				
Boys	-.11 (-.18; -.03)	-.40 (-.47; -.33)	-.21 (-.27; -.14)	-.32 (-.39; -.25)
Girls	-.17 (-.25;-.09)	-.39 (-.45; -.33)	-.19 (-.25; -.13)	-.35 (-.41; -.29)
Environmental Correlation				
Boys	-.01 (-.11; .08)	-.16 (-.26; -.05)	-.05 (-.14; .05)	-.19 (-.28; -.09)
Girls	.01 (-.09; .11)	-.17 (-.27; -.07)	-.05 (-.15; .05)	-.13 (-.23; -.03)

DISCUSSION

The aim of the present study was to determine whether the observed genetic correlation between ODD and ADHD symptoms and educational achievement is best explained by a causal effect of ODD or ADHD symptoms on educational achievement, or by genetic pleiotropy. In line with earlier research we found significant negative associations between ODD and ADHD symptoms and educational achievement (Polderman et al., 2010)(Polderman et al., 2010). Children, who displayed more ODD or ADHD symptoms, as rated by their mother at the same time or 5 years earlier, scored lower on a standardized educational achievement test. Comparing the different components of ADHD, inattentiveness and hyperactivity, suggests variation in the magnitude of the association with educational achievement. Inattentiveness is to a much greater extent related to educational achievement than hyperactivity.

For ODD symptoms the association within genetically identical twin pairs was rather small and only significant for girls and not boys. Moreover, the cross-sectional and longitudinal genetic correlations between ODD symptoms and educational achievement were significant, but most environmental correlations were not. Absence of significant environmental correlations implies that a causal effect of ODD symptoms on educational achievement is falsified and that genetic pleiotropy underlies the association. However, power to detect an environmental correlation was low, a large number of complete twin pairs are necessary when the phenotypic correlation is small, which is the case for the association between ODD symptoms and educational achievement (boys: r = -.08; girls: r = -.09) (de Moor et al, 2008).

Within genetically identical twin pairs, the twin who showed more ADHD symptoms scored lower on the educational achievement test than his or her co-twin. Thus, even when correcting for possible confounding by genes, the association remained significant. The cross-sectional and longitudinal genetic correlations between ADHD symptoms and educational achievement were significant, as were the environmental correlations. This supports the causal effect of ADHD symptoms on educational achievement.

Taken together, the tests do not support a causal effect of ODD symptoms on educational achievement. However, this rejection of the causal hypothesis should be treated with caution as there was a lack of power to detect an environmental correlation due to the small phenotypic correlation that is observed between ODD symptoms and educational achievement (de Moor et al, 2008).

The tests fully supported the causal effect of ADHD symptoms on educational achievement. This indicates that a behavioral intervention or medication prescription, leading to a reduction in symptoms of ADHD (King et al., 2006; Schachter et al., 2001), will also indirectly, through the causal chain, improve the educational achievement of children. The effects of prescription of medication for ADHD on the performance at school have been investigated in earlier research. When medication use resulted in a decrease in symptoms of ADHD, children were indeed better able to stay focused and completed more of their school work (Brown et al., 2005; Prasad et al., 2013). The influence on the actual educational achievement was only modest and evidence was less convincing.

A limitation of this study is that it could not test the direction of the causality and more complex mechanisms of causality, such as bidirectional causality, or a combination of pleiotropy and a reverse causal effect of low educational achievement on ODD and ADHD symptoms. Children who have problems keeping up in school display, perhaps out of frustration, ODD or ADHD symptoms. Bidirectional causality implies that ODD or ADHD symptoms lead to lower educational achievement and in turn problems at school enhance the already existing symptoms. There are direction of causality models that could be used to study these more complex mechanisms (Duffy & Martin, 1994; Heath et al., 1993). However, to be able to resolve the direction of the causal association these models require a substantial difference in heritability, which is not the case for either ODD or ADHD symptoms and educational achievement.

ODD and ADHD symptoms were found to be associated, both cross-sectional and longitudinal, with lower educational achievement in primary school children. The results for ODD symptoms and educational achievement were somewhat inconsistent, probably due to a lack of power, and the causal hypothesis could not be supported. The results for ADHD symptoms and educational achievement were in line with a causal effect of ADHD symptoms on educational achievement. A practical implication following from the casual effect of ADHD symptoms on educational achievement is that, when a behavioral intervention or medication prescription leads to a reduction in ADHD symptoms, it could also have an enhancing influence on educational achievement. This effect will probably be larger for children displaying inattentive symptoms compared to children mainly demonstrating hyperactive symptoms given the difference in the strength of the association with educational achievement.

REFERENCES

American Psychiatric Association (2000). Diagnostic and Statistical Manual of Mental Disorders: 4th ed., text rev. DSM-IV-TR.

Angold, A., Costello, E.J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry, 40 (1),* p. 57-87.

Bartels, M., Rietveld, M.J., van Baal, G.C., & Boomsma, D.I. (2002). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Research, 5 (6),* p. 544-553.

Bartels, M., van Beijsterveldt, C.E., Derks, E.M., Stroet, T.M., Polderman, T.J., Hudziak, J.J., & Boomsma, D.I. (2007). Young Netherlands Twin Register (Y-NTR): a longitudinal multiple informant study of problem behavior. *Twin Research and Human Genetics, 10 (1),* p. 3-11.

Boker S.M., Neale, M.C., Maes, H.H.M., Wilde M.J., Spiegel M., Brick T.R., Spies J., Estabrook R., Kenny S., Bates T.C., Mehta P., & Fox J. (2011). An open source extended structural equation modeling framework. *Psychometrika, 76 (2),* p. 306-317.

Boker, S.M., Neale, M.C., Maes, H.H.M., Wilde, M.J., Spiegel, M., Brick, T.R., Estabrook, R., Bates, T.C., Mehta, P., von Oertzen, T., Gore, R.J., Hunter, M.D., Hackett, D.C., Karch, J., & Brandmaier, A. (2012). OpenMx 1.2 User guide.

Brown, R.T., Amler, R.W., Freeman, W.S., Perrin, J.M., Stein, M.T., Feldman, H.M., Pierce, K., & Wolraich, M.L. (2005). Treatment of attention-deficit/hyperactivity disorder: overview of the evidence. *Pediatrics, 115 (6),* p. e749-e757.

Cito (2002). Eindtoets Basisonderwijs. Arnhem, The Netherlands: Cito.

de Moor, M.H., Boomsma, D.I., Stubbe, J.H., Willemsen, G., & de Geus, E.J. (2008). Testing causality in the association between regular exercise and symptoms of anxiety and depression. *Archives of General Psychiatry, 65 (8),* p. 897-905.

de Zeeuw, E.L., van Beijsterveldt, C.E., Glasner, T.J., Bartels, M., Ehli, E.A., Davies, G.E., Hudziak, J.J., Rietveld, C.A., GroenGÇÉBlokhuis, M.M., & Hottenga, J.J. (2014). Polygenic scores associated with educational attainment in adults predict educational achievement and ADHD symptoms in children. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 165 (6),* p. 510-520.

Duffy, D.L. & Martin, N.G. (1994). Inferring the direction of causation in cross-sectional twin data: theoretical and empirical considerations. *Genetic Epidemiology, 11 (6),* p. 483-502.

Faraone, S.V., Perlis, R.H., Doyle, A.E., Smoller, J.W., Goralnick, J.J., Holmgren, M.A., & Sklar, P. (2005). Molecular genetics of attention-deficit/hyperactivity disorder. *Biological Psychiatry, 57 (11),* p. 1313-1323.

Greven, C.U., Kovas, Y., Willcutt, E.G., Petrill, S.A., & Plomin, R. (2014). Evidence for shared genetic risk between ADHD symptoms and reduced mathematics ability: a twin study. *Journal of Child Psychology and Psychiatry, 55 (1),* p. 39-48.

Haworth, C.M., Asbury, K., Dale, P.S., & Plomin, R. (2011). Added value measures in education show genetic as well as environmental influence. *PloS one, 6 (2)*, p. e16006.
Heath, A.C., Kessler, R.C., Neale, M.C., Hewitt, J.K., Eaves, L.J., & Kendler, K.S. (1993). Testing hypotheses about direction of causation using cross-sectional family data. *Behavior Genetics, 23 (1)*, p. 29-50.
Hudziak, J.J., Derks, E.M., Althoff, R.R., Copeland, W., & Boomsma, D.I. (2005). The genetic and environmental contributions to oppositional defiant behavior: a multi-informant twin study. *Journal of the American Academy of Child and Adolescent Psychiatry, 44 (9)*, p. 907-914.
IBM Corp (2012). SPSS Statistics for Windows Version 21.0. Armonk, NY: IBM Corp.
King, S., Griffin, S., Hodges, Z., Weatherly, H., Asseburg, C., Richardson, G., Golder, S., Taylor, E., Drummond, M., & Riemsma, R. (2006). A systematic review and economic model of the effectiveness and cost-effectiveness of methylphenidate, dexamfetamine and atomoxetine for the treatment of attention deficit hyperactivity disorder in children and adolescents. *Health Technology Assessment, 10 (23)*, p. iii-iiv.
Kuntsi, J., Eley, T.C., Taylor, A., Hughes, C., Asherson, P., Caspi, A., & Moffitt, T.E. (2004). Co-occurrence of ADHD and low IQ has genetic origins. *American Journal of Medical Genetics Part B Neuropsychiatric Genetics, 124B (1)*, p. 41-47.
Lubke, G.H., Hudziak, J.J., Derks, E.M., van Beijsterveldt, T.C., & Boomsma, D.I. (2009). Maternal ratings of attention problems in ADHD: evidence for the existence of a continuum. *Journal of the American Academy of Child and Adolescent Psychiatry, 48 (11)*, p. 1085-1093.
Martin, J., Hamshere, M.L., Stergiakouli, E., O'Donovan, M.C., & Thapar, A. (2014). Neurocognitive abilities in the general population and composite genetic risk scores for attention-deficit hyperactivity disorder. *Journal of Child Psychology and Psychiatry, Epub ahead of print.*
Moilanen, K.L., Shaw, D.S., & Maxwell, K.L. (2010). Developmental cascades: externalizing, internalizing, and academic competence from middle childhood to early adolescence. *Development and psychopathology, 22 (3)*, p. 635-653.
Neale, M.C., Røysamb, E., & Jacobson, K. (2006). Multivariate genetic analysis of sex limitation and GxE interaction. *Twin Research and Human Genetics, 9 (4)*, p. 481-489.
Plomin R., DeFries J.C., McClearn G.E., & McGuffin P.s (2008). Behavioral Genetics. *5th Edition.* New York, NY: Worth Publishers.
Polderman, T.J.C., Boomsma, D.I., Bartels, M., Verhulst, F.C., & Huizink, A.C. (2010). A systematic review of prospective studies on attention problems and academic achievement. *Acta Psychiatrica Scandinavica, 122 (4)*, p. 271-284.
Prasad, V., Brogan, E., Mulvaney, C., Grainge, M., Stanton, W., & Sayal, K. (2013). How effective are drug treatments for children with ADHD at improving on-task behaviour and academic achievement in the school classroom? A

systematic review and meta-analysis. *European Child and Adolescent Psychiatry, 22 (4)*, p. 203-216.

R Core Team (2014). R: A language and environment for statistical computing. Vieanna, Austria: R Foundation for Statistical Computing.

Rietveld, M.J., van, D., V, Bongers, I.L., Stroet, T.M., Slagboom, P.E., & Boomsma, D.I. (2000). Zygosity diagnosis in young twins by parental report. *Twin Research, 3 (3)*, p. 134-141.

Saudino, K.J. & Plomin, R. (2007). Why are hyperactivity and academic achievement related? *Child Development, 78 (3)*, p. 972-986.

Schachter, H.M., King, J., Langford, S., & Moher, D. (2001). How efficacious and safe is short-acting methylphenidate for the treatment of attention-deficit disorder in children and adolescents? A meta-analysis. *Canadian Medical Association Journal, 165 (11)*, p. 1475-1488.

van Beijsterveldt, C.E., Groen-Blokhuis, M., Hottenga, J.J., Franic, S., Hudziak, J.J., Lamb, D., Huppertz, C., de Zeeuw, E., Nivard, M., Schutte, N., Swagerman, S., Glasner, T., van Fulpen, M., Brouwer, C., Stroet, T., Nowotny, D., Ehli, E.A., Davies, G.E., Scheet, P., Orlebeke, J.F., Kan, K.J., Smit, D., Dolan, C.V., Middeldorp, C.M., de Geus, E.J., Bartels, M., & Boomsma, D.I. (2013). The Young Netherlands Twin Register (YNTR): longitudinal twin and family studies in over 70,000 children. *Twin Research and Human Genetics, 16 (1)*, p. 252-267.

9.

SUMMARY AND GENERAL DISCUSSION

SUMMARY

This thesis contributed to knowledge on the causes of individual differences between children in educational achievement by looking at the influence of genetic effects and of twinning, teachers and other environmental factors on educational achievement, as measured by teachers' reports and objective standardized tests. The teacher reports included ratings on arithmetic, language, reading and physical education. The standardized tests included pupil monitoring tests on arithmetic, reading, reading comprehension and spelling for primary school grades (Cito, 2014a; Vlug, 1997) and a national educational achievement test (Cito, 2002) administered in the final grade of primary school, at about 12 years of age, with questions on arithmetic, language, study skills and science and social studies. In addition, we extended the study of individual differences in educational achievement by also looking at their association with problem behaviors, as rated by teachers and mothers. The teacher and mother ratings of problem behavior focused on the presence of symptoms of Oppositional Defiant Disorder (ODD) and Attention Deficit Hyperactivity Disorder (ADHD) and were assessed by the short versions of the Conners' Teacher Rating Scale - Revised and the Conners' Parent Rating Scale - Revised (Conners et al., 1998; Conners, 2001). The data were collected at age 7, 9 and 12 years from mothers and teachers of twins and teachers of non-twin siblings of the twins registered with the Netherlands Twin Register (NTR). I will present a summary of the main findings of each chapter of this thesis and discuss the results, some practical implications and future research.

In the first part of this thesis, several predictors of school performance, educational achievement and problem behavior were examined. **Chapter 2** focused on pre- and perinatal risk factors, more prevalent in twins, for educational achievement. It was established that low birth weight and being small for gestational age were risk factors for lower school performance. These results were robust even after correcting for socioeconomic status (SES). The effects of these risk factors were small, especially when compared to the effect of gender of the child, which had a larger influence. Other pre- and perinatal risk factors that are more prevalent in twins are assisted conception, cesarean section, incubator time and birth complications. With the exception of mode of delivery, these risk factors negatively affected performance in physical education, but they had no other effects on educational achievement. The twin-specific risk factor zygosity had no effect. To test the assumption that twins are not different from singletons, the school performance of twins was compared to that of their non-twin siblings in a within-family design, thereby taking into account confounding of multiple demographic characteristics. There were small differences between twins and singletons in school performance when

comparing the twins to their older siblings, but not when comparing them to their younger siblings. Thus, birth order within the family partly explains the small differences that are often suggested to exist between twins and singletons.

Another potential factor influencing educational achievement is having a teacher of opposite gender, i.e. for boys it is suggested to be detrimental for their performance at school to have a female teacher and girls would benefit from being taught by a female teacher. In **chapter 3**, a contribution is made to the ongoing discussion in society as to whether boys might be disadvantaged by the feminization of primary education. This was done by selecting a subgroup of 12-year old monozygotic and dizygotic of opposite-sex twin pairs where one twin was taught by a male teacher whereas the other twin was taught by a female teacher. As (part of) their genotype, family background, social economic status and multiple other characteristics are more or less similar, differences within the twin pairs may be ascribed to the influence of the gender of the teacher. Boys outperformed girls in arithmetic, while girls scored higher on language and reading. Boys also demonstrated more ADHD related behavior, but these findings were independent of teachers' gender. Therefore, increasing the number of male teachers in primary education or implementing single-gender education may not be as effective to close a possible gender gap in educational achievement or ADHD behavior as suggested by some.

Although there might be some differences between groups of children, e.g. boys and girls, even children attending the same school and taught by the same teacher differ greatly in their performance and behavior at school. In the second part of this thesis, we tried to get a better understanding of why children differ in their educational achievement and problem behavior by exploring the interaction between genetic effects and the environment on educational achievement and problem behavior. The moderating influence of several (environmental) factors on the heritability of educational achievement and problem behavior was explored (Purcell, 2002). Moderation of heritability of educational achievement was investigated for gender and country of the student and moderation of the heritability of ODD and ADHD behavior was investigated for classroom sharing, gender of the student and gender of the teacher. In **chapter 4**, the heritability of educational achievement across several educational domains, i.e. arithmetic, reading, reading comprehension and spelling, across grade 1 to 6 (age 6 to 13), was estimated. Genes explained most of the individual differences in educational achievement across all grades for arithmetic (60-74%), reading (72-82%) and reading comprehension (54-63%). In contrast, the heritability of spelling was smaller in the first grade (33%) compared to later ages (58-70%). The heritability of general educational achievement in the final grade of primary school was high (74%). The common environmental effects had only a small influence on the individual differences in educational achievement. Boys and girls for some educational domains differed in the average test results, but there were no differences between boys and girls in the heritability of educational achievement. This implies that the extent to which genes and the environment influence educational achievement is similar across gender (no quantitative gender differences) and that the genes that have an influence on educational achievement are the same for boys and girls (no qualitative gender differences).

In **chapter 5**, the heritability of educational achievement in the Netherlands is put into perspective by a review of the existing literature on twin studies from different countries estimating heritability of educational achievement in primary school. A PubMed search retrieved 61 studies describing studies from 6 different, mostly English speaking, countries and including subjects from 11 different cohorts. Heritability estimates varied considerably across studies as did the influence of the environmental effects. The small sample sizes, different age groups and the variety of measurement instruments are probably the main reasons for this variability. Meta-analyses of the twin correlations was done to obtain a heritability estimate in the largest sample and test for differences in the heritability between countries. The estimated mean heritability for the educational domains reading (69%), reading comprehension (49%), mathematics (57%) and spelling (44%) and for general educational achievement (66%) was moderate to high. The importance of genetic effects for educational achievement differed between the USA, UK and the Netherlands. The heritability estimates for reading, reading comprehension and mathematics were consistently high in the Netherlands whereas this was not true for the USA and UK, suggesting moderation of the heritability by country. Heritability of reading was equally high across countries (USA: 70%; UK: 69%; NL: 66%), but heritability of reading comprehension was larger in the Netherlands (64%) and the USA (67%) compared to the UK (38%) and heritability of mathematics was low in the USA (26%), moderate in the UK (46%) and high in the Netherlands (71%).

In **chapter 6**, the heritability of ODD and ADHD behavior was estimated at the ages of 7, 9 and 12 years. To this end, it was first tested whether the scales of the short CTRS-R measured the same underlying construct across groups, in other words, whether the scales were measurement invariant (MI) (Millsap & Yun-Tein, 2004). There were two grouping variables, i.e. gender of the teacher and gender of the student, and MI was confirmed for three of the four scales measuring ODD and ADHD behavior, namely oppositional behavior (OPP), hyperactivity (HYP) and ADHD index (ADHD). In contrast, measurement invariance did not hold for the inattention/cognitive problems (ATT) scale.

Even without constraints on the factor structure the model fit was not acceptable for ATT and increasing MI levels resulted in a worsening of the model fit. This strongly questions the reliability of this scale and its use in clinical practice. After having established MI for three scales of the CTRS-R we looked at the extent to which individual differences in ODD and ADHD behavior, as measured by these scales, could be explained by genes and the environment. There were some gender differences in the etiology of ODD behavior. Heritability was higher for boys (OPP - ST: 62-80%; DT: 12-57%) than girls (OPP - ST: 33-46%; DT: 25-55). The ratio between the contribution of additive and non-additive genetic effects resulted in gender differences for the hyperactive component of ADHD behavior (boys - ST: 76-84%; DT: 48-51%; girls - ST: 66-75%; DT: 43-51%). The heritability for ODD and ADHD behavior at school depended in some cases on the gender of the teacher. However, the direction of the effects of gender of the teacher was not consistent across ages and within scales which makes interpreting the findings difficult. Heritability of ODD and ADHD behavior was substantially larger in children who shared a classroom (ST) compared to those who did not (DT) (boys: OPP - ST: 62-80%; DT: 12-57%; girls OPP - ST: 33-46%; DT: 25-55%; boys: HYP - ST: 76-84%; DT: 48-51%; girls: HYP - ST: 66-75%; DT: 43-51%; boys and girls: ADHD - ST: 78-88%; DT: 46-61%). The results excluded teacher bias as an explanation and indicated that the heritability of ODD and ADHD behavior is moderated by the classroom. Apparently the difference in behavior is elicited by different classroom environments, teachers and peers.

In the third part of this thesis the association between ODD and ADHD behavior and educational achievement was investigated by two genetic approaches, namely polygenic score analyses, and causality models. The association of ODD and ADHD with educational achievement is usually negative; children with these problems perform less well in school, and it is an important question what the etiology of this association might be. In **chapter 7** the effect sizes from a large genome wide association (GWA) meta-analysis of educational attainment (Rietveld et al., 2013) were used to calculate polygenic scores in an independent sample of 12-year-old children. This cohort had data on school performance, educational achievement and ADHD symptoms as well as on genome wide single nucleotide polymorphisms (SNPs) available. The polygenic scores explained up to 4, 3 and 2 per cent of the variance in school performance, educational achievement and ADHD behavior, respectively. Clearly, some of the genetic variants that influence educational attainment in adults also have an effect on school performance and educational achievement in children. Moreover, the genetic variants also had a significant effect on

ADHD behavior. This confirms at the measured genotype level a genetic association between educational achievement and ADHD.

The question that remains is whether this genetic correlation is due to genetic pleiotropy, where the same genetic variants influence multiple (brain) phenotypes, or due to a causal effect of ADHD on educational achievement. The genetic variants causing ADHD will then indirectly also cause low educational achievement. The causal hypothesis of a detrimental effect of ODD and ADHD behavior on educational achievement has high face validity: being genetically predisposed to ADHD, for instance, could make it harder to concentrate at school, leading to lower educational achievement. In **chapter 8** the hypothesis of a causal effect was tested against the null hypothesis of genetic pleiotropy in a large genetically sensitive sample. Children who displayed more ODD or ADHD behavior scored lower on educational achievement and this was true for ODD and ADHD behavior measured at the same age as well as 5 years earlier. The results suggest a causal effect as most likely explanation for the association between ADHD and educational achievement. First, in genetically identical twins, the child who shows more ADHD behavior than his co-twin also performs worse at school. Thus within genetically identical twin pairs, a design correcting for possible genetic confounding, there was a negative association between ADHD behavior and educational achievement. Second, both the cross-sectional and longitudinal genetic and environmental correlation between ADHD and educational achievement were significant. The genetic as well as the environmental effects with an influence on ADHD behavior also affected educational achievement, supporting the causal effect hypothesis. The results for ODD behavior and educational achievement were less consistent, due to a lack of power, and only partly supported a causal effect. Thus, the causal effect for ODD behavior on educational achievement could not be falsified.

GENERAL DISCUSSION

In the last decade (2003-2013) the number of twin births in the Netherlands decreased from 18.3 to 16.5 per 1.000 births (Wobma & Garssen, 2014). This decline in twin births is, in general, seen as a positive trend since twin births are often associated with a higher prevalence of risk factors during pregnancy and birth (Glasner et al., 2012). These risk factors include prematurity, low birth weight and birth complications and have a higher prevalence in twins than in singleton births. They all tend to be associated with negative health outcomes and possibly also with a negative influence on educational achievement (Khadem & Khadivzadeh, 2010; Lundgren & Tuvemo, 2008; Wagenaar et al., 2008). In this thesis we describe that low birth weight and being small for gestational age were indeed relevant risk factors for educational achievement, but in comparison to the effects of socioeconomic status (SES) and gender, their effects were rather small.

The average difference in birth weight between twins and singletons is more than 1000 grams (De Geus et al., 2001) and the gestational age of twins is on average 3-4 weeks shorter than that of singletons (Gielen et al., 2010). This raises the question whether these pre- and perinatal risk factors, which are more prevalent in twins, might explain the differences between twins and their non-twin siblings that often have been suggested for general cognitive ability and educational achievement. The results of **chapter 2** indicate that, in 7-year-olds, there are small differences in educational achievement between twins and their non-twin siblings. However, when taking into account the birth order within the family, the differences between twins and siblings disappear. Because twins often are the last born children in a family, not taking into account birth order - even when using an optimal design that compares twins to their own siblings- might lead to the wrong conclusion that 'being a twin' is a risk factor for lower educational achievement. This is an important finding, also suggesting that pre- and perinatal risk factors that are more prevalent in twins do not lead to long term twin-sibling differences and it is supported by observations that the small observed differences due to birth order dissipate when the twins grow older. From research in other domains, such as general cognitive ability (Webbink et al., 2008), body composition (Estourgie-van Burk et al., 2010) and development of ADHD symptoms (Robbers et al., 2011), it is known that the differences between twins and singletons disappeared at later ages and that twins do not seem to differ from singletons in educational achievement, behavior or health (Petersen et al., 2011) and twin data are a valuable resource to draw conclusions about heritability that may be generalized to the population at large.

A further consideration is whether the twin pairs in the study sample from the NTR were representative of the general population of twins in the Netherlands. Exclusion criteria for analyses of phenotypes reported on in this thesis by teachers were a disease or handicap that interfered severely with daily functioning, and attendance of specialized education. In the Dutch educational system, special schools are available for children who need extra care due to learning problems, physical and/or mental disabilities or behavioral disorders. This means that the lower end of the distribution was not represented in the teacher sample for both twins and singletons. A bias would be present when more twins are referred to specialized education compared to non-twin singleton children. To our knowledge, there are no national statistics available on the percentage of multiples attending specialized education. As indicated before a large percentage of children are part of a multiple (Wobma & Garssen,

2014) and it is astonishing that no data are available on the total number of twins in specialized education. In our sample, parents reported for approximately 4 per cent (age 7: 2.9%; age 9: 5.2%; age 12: 5.3%) of the twins that they attended specialized education while in the Netherlands around 5 per cent of all school aged children attend some type of specialized education (CBS Statistics Nederland, 2014). However, this lower percentage in NTR twins may not accurately reflect the percentage of Dutch twins in specialized education. Parents might have more often decided to refrain from participating in research of the NTR when one of their children is a child with special needs.

With a unique design of identical twin pairs discordant for the gender of their teacher and dizygotic twins of opposite-sex concordant for the gender of their teacher, we made a contribution to the ongoing debate in the media and society about the declining number of male teachers in the educational system and its negative effect on the performance and behavior of boys in school. Some people argue that a same-gender teacher enhances the performance of a child at school because students identify themselves more with a same-gender teacher (Carrington, Tymms & Merrell, 2008), teachers feel more competent with a same-gender student (Powell & Downey, 1997) or by the effects of stereotype threat (Steele, 1997). However, we found no evidence for an effect of a same-gender teacher on educational achievement or ADHD behavior.

An underlying issue in this debate is whether boys are actually underperforming at school compared to girls ('boys problem') (Ailwood, 2003; Carrington, Tymms & Merrell, 2008; van Langen & Driessen, 2006). The existence of an overall lower performance for boys compared to girls has not been found in our primary school sample. However, substantial traditional gender differences were observed with boys scoring higher on numeracy domains and girls performing better on literacy domains. Boys received higher teacher ratings for arithmetic and performed much better on the standardized tests for arithmetic whereas girls received higher teacher ratings for language and reading and performed somewhat better on the standardized tests for language and reading comprehension. These gender differences are also present in the recent national cohorts of Dutch primary school children participating in the educational achievement test administered in the final grade (Cito, 2014b). The percentage of boys, in the NTR data, scoring in the highest category of the total score of the educational achievement test was somewhat higher than the percentage of girls. It should be noted that boys are more likely to have to repeat a grade and are more often attending specialized education with a ratio for boys and girls of approximately 2.5:1 (CBS Statistics Nederland, 2014). This suggests that boys are overrepresented among the underperformers as well as the high performers.

Mean differences between boys and girls explain part of the variance in educational achievement between children, but even after taking these gender differences into account, there are large differences between children in their educational achievement. The underlying causes, genes or the environment, of these individual differences in educational achievement between children are the same for boys and girls as was shown in **chapter 4**. Heritability of educational achievement is substantial and relatively stable across all grades of primary school in the Netherlands in both genders. Genes are the most important cause of individual differences between children in their educational achievement for the core educational domains, i.e. arithmetic, reading (comprehension) and spelling. This contrasts with general cognitive ability where, in children, the environment explains the largest part of the individual differences (Van Soelen et al., 2011). The heritability of general cognitive ability increases significantly and linearly from 41% in childhood (9 years) to 55% in adolescence (12 years) and to 66% in young adulthood (17 years), as demonstrated in a sample of 11,000 pairs of twins from four countries (Haworth et al., 2010). In the NTR, the heritability of general cognitive ability at younger ages is estimated even lower (Bartels et al., 2002).

Often general cognitive ability is thought to be an 'innate' ability while educational achievement is seen as the result of several factors, including but not limited to general cognitive ability. Hence, it seems counterintuitive that the heritability of educational achievement is higher than the heritability of general cognitive ability. One hypothesis for this difference is that the homogeneity of education reduces differences in the environment and, as a result, individual differences between children in educational achievement can to a greater extent be explained by genes (Heath et al., 1985). Studies in preschool children report a much larger influence of the common environment, shared by all children in a family, on, for example, reading (Byrne et al., 2009; Oliver, Dale & Plomin, 2005), than has been found for school going children (Kovas et al., 2013). It could be that the common environment for educational achievement mainly consists of the educational system and school environment, whereas the common environment that influences general cognitive ability has many more aspects. Obviously, ' common environment' for twins and siblings can only be 'common' when twins go to the same class or school and common environment for siblings of different ages also will be less, as they nearly always attend different classes. Homogeneity of the school environment and educational system would reduce the impact of the common environment on educational achievement but not necessarily on general cognitive ability. The influence of the common environment on general cognitive ability clearly decreases when children grow up (Bartels et al., 2002; Haworth et al., 2011). An influential hypothesis states that one reason is that children increasingly have the opportunity to select their own unique environments when they grow up (Deary et al., 2012; Molenaar et al., 2013).

Although heritability of educational achievement is high in most Western societies, which is reflected in the overview presented in **chapter 5**, there are some differences between countries. In the Netherlands, heritability was consistently high across different educational domains whereas the variability in estimates was larger across different educational domains for the USA and UK. This is an indication of moderation of the heritability by country. An explanation might be that the equality in income and circumstances under which children grow up, but importantly also the heterogeneity in educational opportunity, is larger in the Netherlands compared to the USA and the UK.

The consequence of the homogeneity in an educational system is that it will highlight the innate individual differences between children as reflected in the high heritability (Harlaar et al., 2012; Kovas et al., 2013). What must be kept in mind is that heritability does not equal determinism. The variance between children may be heritable, but the mean of a population or a group can be positively influenced by a good quality school environment. High heritability in a homogeneous school environment can imply that children with a predisposition for lower educational achievement will have to struggle while children with a genetic advantage can excel at school without ever tapping their full potential. High heritability therefore supports the importance of differentiation in teaching. The double challenge for primary school teachers is to make sure that children, who have more difficulty at school, will learn reading, writing and arithmetic, but that those who have it easy are still sufficiently challenged. Classroom teaching might not be the best method to achieve this goal and a more personalized approach to learning may be warranted. Unfortunately, the increasing number of children per teacher and the demand on teachers with regard to administrative duties might preclude teachers from customizing their lessons to the needs of each child.

Some parts of a child's environment that we regard as 'common', or shared by children from the same family, like parental educational level, are influenced by parental genotype (Rietveld et al., 2013; Vinkhuyzen et al., 2010). Because the children share genes with their parents, the genes of a child can become correlated with its environment, i.e. passive gene-environment (GxE) correlation. The common environment did not seem to have much of an influence on educational achievement in children when correcting for this genetic confounding. However, there are several mechanisms through which the common environment, e.g. SES and parental upbringing, can still influence educational achievement. The common environment can have an influence through gene-environment interaction, thereby having a different influence on siblings who have different genotypes. Also, the influence of the environment may not be uniform across the entire distribution, for example, the influence of the environment appears to be larger in so called high-risk home environments (low SES) while it has no influence in more advantaged homes (Scarr-Salapatek, 1971; Turkheimer et al., 2003). It could also be that the impact of the common environment on a child's educational achievement may be at a child-specific level rather than at a family-wide level (McGue & Bouchard Jr, 1998). The interaction with the teacher at school, for example, may be experienced rather differently by children from the same family, transforming these common environmental factors into unique environmental effects (Somersalo, Solantaus & Almqvist, 2002).

A favorable role for factors like SES and parental upbringing has been shown in adoption studies were children grow up in the same environment with parents that they are not genetically related to. Adoptive families mostly have a SES above average while the opposite is true for the biological families. Adoption studies found that general cognitive ability and educational achievement of adopted children was higher than of their non-adopted biological siblings, who were raised by their birth parents (Maughan, Collishaw & Pickles, 1998; Scarr & Weinberg, 1983; Van IJzendoorn, Juffer & Poelhuis, 2005). However, the scores for the adopted children were lower than those for the biological children of their adoptive parents and individual differences among them were more related to differences among their biological than adoptive parents, whether they lived together or not. Young siblings were found to be quite similar, whether genetically related or not, but adolescents' general cognitive ability scores were similar to those of their parents and siblings only if they were biologically related (Scarr & Weinberg, 1983).

Educational achievement is known to relate to childhood problem behaviors, including ODD and ADHD. Expression of these behaviors appears to be sensitive to the classroom setting. The heritability estimates of both ODD and ADHD behavior were much larger in children sharing a classroom compared to children in different classrooms. Different classrooms with different peers, teachers and classroom settings trigger different behavior in children depending on their genotype (Eaves, 1984). A teacher might be a very important factor in the expression of the problem behavior in a child with a predisposition for ODD or ADHD. Teachers differ in, for example, the structure of their teaching and the rules children have to comply with in the classroom. When a child displays, for example, ADHD behavior in the classroom, which is to a large extent genetically influenced, a teacher will respond to the behavior of this child in his or her own way. As a consequence the behavior of the teacher changes the environment which can then have an influence on the child's expression of the genes associated with ADHD behavior. These findings are clinically relevant as it implies the possibility for the school to implement teacher based interventions to buffer against the genetic vulnerability of developing ODD or ADHD (Reinke & Herman, 2002).

Raters can have their own perception on behavior which makes ratings by the same person of multiple children more similar. Although this type of rater bias was ruled out as an explanation, the lower heritability of ODD and ADHD behavior for children in different classrooms may be the result of other rater effects. It has been demonstrated that, when rater specific factors are genetically influenced, heritability estimates depend on whether children are assessed by the same rater or by two different raters (Kan et al., 2014). When heritability estimates for ADHD behavior are based on ratings from different teachers only the component of the phenotype that they both agree on, the common component, is estimated while the rater specific component ends up in the environmental effects (McLoughlin et al., 2011; Merwood et al., 2013). When the same teacher rates both members of a twin pair the component that only the teacher observes, the rater specific component, will contribute to the genetic effects which results in a higher estimated heritability.

The interpretation of the heritability in the presence of rater specific genetic factors requires caution, but the existence of rater specific factors is not inconsistent with GxE interaction. The more environments differ the less genetic variance tends to be shared between raters observing a child in the different classroom environments. This also explains the finding that the genetic overlap between mother and father ratings is larger than between parent and teacher ratings (Merwood et al., 2013). In primary school, children are sometimes taught by two teachers, for example when teachers do not work fulltime, as is increasingly common in the Netherlands. Ratings from two teachers each rating both children from a twin pair sharing a classroom could improve our understanding of GxE interaction and the contribution of rater specific genetic factors to ODD and ADHD behavior.

Chapter 6 reports that for ADHD, additive as well as non-additive (dominant) genetic effects were relevant for the differences between children. ADHD seems to be influenced by interactions between genes and is not just a summation of the effects of the different genes. Such strong evidence for the influence of interacting genes is hardly ever seen for other behavioral traits or disorders (Burt, 2009). This is somewhat surprising given the large comorbidity between ADHD and other disorders, e.g. externalizing behavior (Angold, Costello & Erkanli, 1999). To examine this difference in genetic architecture between

ADHD behavior and other childhood psychopathologies is an important future research step.

A clinically relevant finding from **chapter 8** is that the negative association between ADHD behavior and educational achievement is likely to be due to a causal effect of ADHD behavior on educational achievement instead of genetic pleiotropy, where the same genetic variants have an effect on both ADHD behavior and educational achievement. This indicates that a behavioral intervention or medication prescription, leading to a reduction in symptoms of ADHD (King et al., 2006; Schachter et al., 2001), will also indirectly, through the causal chain, improve the educational achievement of children. The effects of prescription of medication for ADHD on the performance at school have been investigated in earlier research. When medication use resulted in a decrease in symptoms of ADHD, children were indeed better able to stay focused and completed more of their school work (Brown et al., 2005; Prasad et al., 2013). The influence on the actual educational achievement was only modest and evidence was less convincing.

This outcome is independent of the current discussion on whether children are nowadays too often diagnosed with ADHD and prescribed medication. It is well recognized that an ADHD diagnosis can be considered the extreme end of the normal distribution of inattentive and hyperactive symptoms in the population (Groen-Blokhuis et al., 2014; Lubke et al., 2009). However, it seems that the perception of what is normal for a child of a certain age has changed over the years. Data from the NTR indicate that the number of symptoms associated with ADHD, as reported by parents, has remained very similar over a period of 25 years. If more children have received an ADHD diagnosis and are prescribed medication, there are other reasons beyond the number of symptoms that contribute to such an increase.

A next important next step towards understanding the underlying causes for individual differences between children in behavior and educational achievement would be to investigate the causes of continuity and change. There are two main hypotheses with regard to the underlying mechanisms of development (Rowe & Britt, 1991). A transmission model assumes that educational achievement at different grades is causally linked and preceding experiences are transmitted to later time points. A liability model assumes a stable underlying liability which explains the association between educational achievement in different grades. In a longitudinal twin study across all grades of primary school, these models can be studied at the genetic and environmental level. Genetic and environmental effects might exert their influence on educational achievement through the same or different developmental models. The mechanisms underlying continuity are especially important as the longer a child is performing below average at school, the more difficult it will be to prevent that child from falling behind. The same is true for ODD and ADHD behavior since the longer a child deviates from normal development, the more difficult it will be to successfully intervene and put it back on a normal developmental trajectory (Sroufe, 1990).

Earlier research towards continuity and change in behavioral problems has found that genetic effects are partly transmitted to later ages with some new genes coming into play at each age, whereas the influence of common environmental effects remains the same across development and unique environmental effects are important but specific to a certain age (Bartels et al., 2004). Apparently, for internalizing and externalizing problems, the unique environment is mainly specific to a certain age and has no long-term effect whereas the common environmental effects persist over time. Studies towards developmental trajectories and underlying causes of stability and change are still lacking for educational achievement. NTR started to collect teacher surveys at the ages 7, 9 and 12 years. The longitudinal data collection is now providing a unique opportunity to investigate the stability of ODD and ADHD behavior and educational achievement without the bias of a constant rater, since the longitudinal data on the behavior and school performance of a child come from different teachers at different ages and from objective tests.

Another important next step towards understanding the underlying causes for individual differences between children in behavior and educational achievement would be to identify genetic variants, and their biological mechanisms, related to educational achievement and ODD and ADHD behavior. Knowledge of these causes could lead to more effective interventions and the development of preventions. Heritability of educational achievement and of ODD and ADHD behavior is relatively high, but the identification of genetic variants associated with educational achievement or with a risk for ODD or ADHD has turned out to be much harder than expected. Genome wide association (GWA) studies have been less successful for phenotypes, such as ADHD (Neale et al., 2010) and educational achievement (Rietveld et al., 2013), than for physical phenotypes, such as height (Wood et al., 2014). One of the explanations is that behavioral phenotypes are highly complex and caused by many genetic variants with only small effects (Flint & Munafo, 2013). Other explanations are that it is more difficult to obtain large sample sizes for behavioral phenotypes and that different definitions or assessments of the phenotype might lead to heterogeneity which makes gene finding also more difficult (Wray & Maier, 2014).

Polygenic scores established that a large number of common genetic variants are contributing to variance in the phenotype of interest (Purcell et al., 2009).

The effects of these variants are not significant if tested against genome-wide significance thresholds, but polygenic score analysis allows for an exploration of the underlying etiology of an association between two phenotypes at the measured genotype level, because the information from genome-wide studies is combined into a weighted score across a large number of variants. In **chapter 7**, polygenic scores indicated that genetic variants associated with educational attainment in adults also had an influence on ADHD behavior in children. This demonstrates that even though only a few genetic variants are associated with a phenotype at the genome-wide significance threshold, the information from GWA studies can be used to further our understanding of the causes of the association between two phenotypes.

Single measured genetic variants will likely not be used in the near future to predict a predisposition for ODD or ADHD or for lower educational achievement. However, the authors of the GWA meta-analysis of educational attainment in adults calculated the power of polygenic scores as a prediction variable to make decisions regarding which children to target for interventions, e.g. pre-school programs, when sample sizes will continue to increase. A sample size of 500.000 individuals will be large enough to obtain polygenic scores that can explain 12 per cent of the variance in educational achievement, when other predictors already explain 10 per cent of the variance. This can result in a reduction of 13 per cent in sample size for an intervention and, as the costs of genotyping continue to drop, reducing the number of children to be included in costly interventions could result in substantial savings (Rietveld et al., 2013).

Identification of environmental factors does not seem to be much easier than finding genetic variants responsible for the variance between children in educational achievement and behavior. The unique environmental factors investigated so far do not account for a substantial proportion of the individual differences between children. Nonetheless, in a meta-analysis the proportion of the total variance in, amongst others, behavioral problems and general cognitive ability, explained by differences in peer interaction and student-teacher interaction, reached up to 5 per cent, which is considerably larger than the effect sizes found for genetic variants (Turkheimer & Waldron, 2000). Common environmental factors might explain larger proportions of variance and may therefore be easier to identify, but suffer as indicated earlier, from the difficulty that what we conceive of as common environment, reflects parental genotype (e.g. SES and parental education).

The co-twin control design is an attractive research design to find an association between an environmental factor and a phenotype and to test for causality of this association. This method corrects for several confounding factors, including genetic effects and several other environmental effects, such as shared SES. For example, a study including discordant twins revealed that the negative association between low birth weight and attention problems was probably due to a causal effect (Groen-Blokhuis et al., 2011) Another discordant twin study, disproved anesthesia as an environmental factor with a negative effect on educational achievement and general cognitive ability (Bartels, Althoff & Boomsma, 2009). Similarly, a study on differences between monozygotic twins in their perception of the classroom environment identified the perception of a student of the relationship with the teacher as a unique environmental factor that differed between the genetically identical twins and was linked to hyperactivity as rated by the teacher (Somersalo, Solantaus & Almqvist, 2002).

Further exploration of genetic variants, in larger sample sizes, and of environmental factors, using the methods that control for the possible confounding by genetic effects, will hopefully point to genuine causal associations making it possible to develop new prevention programs and interventions to ensure that each child masters the basic skills at school necessary to succeed in society.

REFERENCES

Ailwood, J. (2003). A national approach to gender equity policy in Australia: another ending, another opening? *International Journal of Inclusive Education, 7 (1)*, p. 19-32.

Angold, A., Costello, E.J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry, 40 (1)*, p. 57-87.

Bartels, M., Rietveld, M.J., Van Baal, G.C., & Boomsma, D.I. (2002). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Research, 5 (6)*, p. 544-553.

Bartels, M., Van den Oord, E.J.C.G., Hudziak, J.J., Rietveld, M.J.H., Van Beijsterveldt, C.E.M., & Boomsma, D.I. (2004). Genetic and environmental mechanisms underlying stability and change in problem behaviors at ages 3, 7, 10, and 12. *Developmental Psychology, 40 (5)*, p. 852-867.

Bartels, M., Althoff, R.R., & Boomsma, D.I. (2009). Anesthesia and cognitive performance in children: no evidence for a causal relationship. *Twin Research and Human Genetics, 12 (3)*, p. 246-253.

Brown, R.T., Amler, R.W., Freeman, W.S., Perrin, J.M., Stein, M.T., Feldman, H.M., Pierce, K., & Wolraich, M.L. (2005). Treatment of attention-deficit/hyperactivity disorder: overview of the evidence. *Pediatrics, 115 (6)*, p. e749-e757.

Burt, S.A. (2009). Rethinking environmental contributions to child and adolescent psychopathology: a meta-analysis of shared environmental influences. *Psychological Bulletin, 135 (4)*, p. 608-637.

Byrne, B., Coventry, W.L., Olson, R.K., Samuelsson, S., Corley, R., Willcutt, E.G., Wadsworth, S., & DeFries, J.C. (2009). Genetic and environmental influences on aspects of literacy and language in early childhood: continuity and change from preschool to Grade 2. *Journal of Neurolinguistics, 22 (3)*, p. 219-236.

Carrington, B., Tymms, P., & Merrell, C. (2008). Role models, school improvement and the gender gap: do men bring out the best in boys and women the best in girls? *British Educational Research Journal, 34 (3)*, p. 315-327.

CBS Statistics Nederland (2014). Speciale scholen: leerlingen, schooltype, leeftijd.

Cito (2002). Eindtoets Basisonderwijs. Arnhem, The Netherlands: Cito.

Cito (2014a). Cito Leerling Volgsysteem. Arnhem, The Netherlands: Cito.

Cito (2014b). Terugblik en resultaten 2014 - Eindtoets Basisonderwijs. Arnhem, The Netherlands: Cito.

Conners, C.K. (2001). Conners' rating scales - revised. New York, NY: Multi-Health Systems, Inc.

Conners, C.K., Sitarenios, G., Parker, J.D., & Epstein, J.N. (1998). Revision and restandardization of the Conners Teacher Rating Scale (CTRS-R): factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology, 26 (4)*, p. 279-291.

De Geus, E.J., Posthuma, D., IJzerman, R.G., & Boomsma, D.I. (2001). Comparing blood pressure of twins and their singleton siblings: being a twin does not affect adult blood pressure. *Twin Research, 4 (5),* p. 385-391.

Deary, I.J., Yang, J., Davies, G., Harris, S.E., Tenesa, A., Liewald, D., Luciano, M., Lopez, L.M., Gow, A.J., Corley, J., Redmond, P., Fox, H.C., Rowe, S.J., Haggarty, P., McNeill, G., Goddard, M.E., Porteous, D.J., Whalley, L.J., Starr, J.M., & Visscher, P.M. (2012). Genetic contributions to stability and change in intelligence from childhood to old age. *Nature, 482 (7384),* p. 212-215.

Eaves, J.L. (1984). The resolution of genotype x environment interaction in segregation analysis of nuclear families. *Genetic Epidemiology, 1,* p. 215-228.

Estourgie-van Burk, G.F., Bartels, M., Boomsma, D.I., & Delemarre-van de Waal, H. (2010). Body size of twins compared with siblings and the general population: from birth to late adolescence. *The Journal of Pediatrics, 156 (4),* p. 586-591.

Flint, J., & Munafo, M. (2013). Genetics. Herit-ability. *Science, 340 (6139),* p. 1416-1417.

Gielen, M., van Beijsterveldt, C.E., Derom, C., Vlietinck, R., Nijhuis, J.G., Zeegers, M.P., & Boomsma, D.I. (2010). Secular trends in gestational age and birthweight in twins. *Human Reproduction, 25 (9),* p. 2346-2353.

Glasner, T.J., van Beijsterveldt, C.E., Willemsen, G., & Boomsma, D.I. (2012). Multiple births in the Netherlands. *Nederlands Tijdschrift voor Geneeskunde, 157 (A5962),* p. 1-5.

Groen-Blokhuis, M.M., Middeldorp, C.M., Kan, K.J., Abdellaoui, A., van Beijsterveldt, C.E., Ehli, E.A., Davies, G.E., Scheet, P.A., Xiao, X., & Hudziak, J.J. (2014). Attention-deficit/hyperactivity disorder polygenic risk scores predict attention problems in a population-based sample of children. *Journal of the American Academy of Child and Adolescent Psychiatry, 53 (10),* p. 1123-1129.

Groen-Blokhuis, M.M., Middeldorp, C.M., van Beijsterveldt, C.E., & Boomsma, D.I. (2011). Evidence for a causal association of low birth weight and attention problems. *Journal of the American Academy of Child and Adolescent Psychiatry, 50 (12),* p. 1247-1254.

Harlaar, N., Kovas, Y., Dale, P.S., Petrill, S.A., & Plomin, R. (2012). Mathematics is differentially related to reading comprehension and word decoding: evidence from a genetically sensitive design. *Journal of Educational Psychology, 104 (3),* p. 622.

Haworth, C.M., Wright, M.J., Luciano, M., Martin, N.G., de Geus, E.J., van Beijsterveldt, C.E., Bartels, M., Posthuma, D., Boomsma, D.I., Davis, O.S., Kovas, Y., Corley, R.P., Defries, J.C., Hewitt, J.K., Olson, R.K., Rhea, S.A., Wadsworth, S.J., Iacono, W.G., McGue, M., Thompson, L.A., Hart, S.A., Petrill, S.A., Lubinski, D., & Plomin, R. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry, 15 (11),* p. 1112-1120.

Haworth, C.M., Asbury, K., Dale, P.S., & Plomin, R. (2011). Added value measures in education show genetic as well as environmental influence. *PloS One, 6 (2),* p. e16006

Heath, A.C., Berg, K., Eaves, L.J., Solaas, M.H., Corey, L.A., Sundet, J., Magnus, P., & Nance, W.E. (1985). Education policy and the heritability of educational attainment. *Nature, 314 (6013),* p. 734-736.

Kan, K.J., van Beijsterveldt, C.E., Bartels, M., & Boomsma, D.I. (2014). Assessing genetic influences on behavior: informant and context dependency as illustrated by the analysis of attention problems. *Behavior Genetics, 44 (4),* p. 326-336.

Khadem, N., & Khadivzadeh, T. (2010). The intelligence quotient of school aged children delivered by cesarean section and vaginal delivery. *Iranian Journal of Nursing and Midwifery Research, 15 (3)*, p. 135-140.

King, S., Griffin, S., Hodges, Z., Weatherly, H., Asseburg, C., Richardson, G., Golder, S., Taylor, E., Drummond, M., & Riemsma, R. (2006). A systematic review and economic model of the effectiveness and cost-effectiveness of methylphenidate, dexamfetamine and atomoxetine for the treatment of attention deficit hyperactivity disorder in children and adolescents. *Health Technology Assessment, 10 (23)*, p. iii-iiv.

Kovas, Y., Voronin, I., Kaydalov, A., Malykh, S.B., Dale, P.S., & Plomin, R. (2013). Literacy and numeracy are more heritable than intelligence in primary school. *Psychological Science, 24 (10)*, p. 2048-2056.

Lubke, G.H., Hudziak, J.J., Derks, E.M., van Bijsterveldt, T.C., & Boomsma, D.I. (2009). Maternal ratings of attention problems in ADHD: evidence for the existence of a continuum. *Journal of the American Academy of Child & Adolescent Psychiatry, 48 (11)*, p. 1085-1093.

Lundgren, E.M., & Tuvemo, T. (2008). Effects of being born small for gestational age on long-term intellectual performance. *Best Practice and Research: Clinincal Endocrinology and Metabolism, 22 (3)*, p. 477-488.

Maughan, B., Collishaw, S., & Pickles, A. (1998). School achievement and adult qualifications among adoptees: a longitudinal study. *Journal of Child Psychology and Psychiatry, 39 (5)*, p. 669-685.

McGue, M., & Bouchard Jr, T.J. (1998). Genetic and environmental influences on human behavioral differences. *Annual Review of Neuroscience, 21 (1)*, p. 1-24.

McLoughlin, G., Rijsdijk, F., Asherson, P., & Kuntsi, J. (2011). Parents and teachers make different contributions to a shared perspective on hyperactive-impulsive and inattentive symptoms: a multivariate analysis of parent and teacher ratings on the symptom domains of ADHD. *Behavior Genetics, 41 (5)*, p. 668-679.

Merwood, A., Greven, C.U., Price, T.S., Rijsdijk, F., Kuntsi, J., McLoughlin, G., Larsson, H., & Asherson, P.J. (2013). Different heritabilities but shared etiological influences for parent, teacher and self-ratings of ADHD symptoms: an adolescent twin study. *Psychological Medicine, 43 (9),* p. 1973-1984.

Millsap, R.E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research, 39 (3)*, p. 479-515.

Molenaar, D., van der Sluis, S., Boomsma, D.I., Haworth, C.M., Hewitt, J.K., Martin, N.G., Plomin, R., Wright, M.J., & Dolan, C.V. (2013). Genotype by environment interactions in cognitive ability: a survey of 14 studies from four countries covering four age groups. *Behavior Genetics, 43 (3),* p. 208-219.

Neale, B.M., Medland, S.E., Ripke, S., Asherson, P., Franke, B., Lesch, K.P., Faraone, S.V., Nguyen, T.T., Schafer, H., Holmans, P., Daly, M., Steinhausen, H.C., Freitag, C., Reif, A., Renner, T.J., Romanos, M., Romanos, J., Walitza, S., Warnke, A., Meyer, J., Palmason, H., Buitelaar, J., Vasquez, A.A., Lambregts-Rommelse, N., Gill, M., Anney, R.J., Langely, K., O'Donovan, M., Williams, N., Owen, M., Thapar, A., Kent, L., Sergeant, J., Roeyers, H., Mick, E., Biederman, J., Doyle, A., Smalley, S., Loo, S., Hakonarson, H., Elia, J., Todorov, A., Miranda, A., Mulas, F., Ebstein, R.P., Rothenberger, A., Banaschewski, T., Oades, R.D., Sonuga-Barke, E., McGough, J., Nisenbaum, L., Middleton, F., Hu, X., & Nelson, S. (2010). Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. . *Journal of the American Academy of Child and Adolescent Psychiatry, 49 (9),* p. 884-897.

Oliver, B.R., Dale, P.S., & Plomin, R. (2005). Predicting literacy at age 7 from preliteracy at age 4: a longitudinal genetic analysis. *Psychological Science, 16 (11),* p. 861-865.

Petersen, I., Nielsen, M.M.F., Beck-Nielsen, H., & Christensen, K. (2011). No evidence of a higher 10 year period prevalence of diabetes among 77,885 twins compared with 215,264 singletons from the Danish birth cohorts 1910–1989. *Diabetologia, 54 (8),* p. 2016-2024.

Powell, B., & Downey, D.B. (1997). Living in single-parent households: an investigation of the same-sex hypothesis. *American Sociological Review, 62 (4),* p. 521-539.

Prasad, V., Brogan, E., Mulvaney, C., Grainge, M., Stanton, W., & Sayal, K. (2013). How effective are drug treatments for children with ADHD at improving on-task behaviour and academic achievement in the school classroom? A systematic review and meta-analysis. *European Child and Adolescent Psychiatry, 22 (4),* p. 203-216.

Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research, 5 (6),* p. 554-571.

Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., & Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature, 460 (7256),* p. 748-752.

Reinke, W.M., & Herman, K.C. (2002). Creating school environments that deter antisocial behaviors in youth. *Psychology in the Schools, 39 (5),* p. 549-559.

Rietveld, C.A., Medland, S.E., Derringer, J., Yang, J., Esko, T., Martin, N.W., Westra, H.J., Shakhbazov, K., Abdellaoui, A., Agrawal, A., Albrecht, E., Alizadeh, B.Z., Amin, N., Barnard, J., Baumeister, S.E., Benke, K.S., Bielak, L.F., Boatman, J.A., Boyle, P.A., Davies, G., de, L.C., Eklund, N., Evans, D.S., Ferhmann, R., Fischer, K., Gieger, C., Gjessing, H.K., Hagg, S., Harris, J.R., Hayward, C., Holzapfel, C., Ibrahim-Verbaas, C.A., Ingelsson, E., Jacobsson, B., Joshi, P.K., Jugessur, A., Kaakinen, M., Kanoni, S., Karjalainen, J., Kolcic, I., Kristiansson, K., Kutalik, Z., Lahti, J., Lee, S.H., Lin, P., Lind, P.A., Liu, Y., Lohman, K., Loitfelder, M., McMahon, G., Vidal, P.M., Meirelles, O., Milani, L., Myhre, R., Nuotio, M.L., Oldmeadow, C.J., Petrovic, K.E., Peyrot, W.J., Polasek, O., Quaye, L., Reinmaa, E., Rice, J.P., Rizzi, T.S., Schmidt, H., Schmidt, R., Smith, A.V., Smith, J.A., Tanaka, T., Terracciano, A., van der Loos, M.J., Vitart, V., Volzke, H., Wellmann, J., Yu, L., Zhao, W., Allik, J., Attia, J.R., Bandinelli, S., Bastardot, F., Beauchamp, J., Bennett, D.A., Berger, K., Bierut, L.J., Boomsma, D.I., Bultmann, U., Campbell, H., Chabris, C.F., Cherkas, L., Chung, M.K., Cucca, F., de, A.M., De Jager, P.L., De Neve, J.E., Deary, I.J., Dedoussis, G.V., Deloukas, P., Dimitriou, M., Eiriksdottir, G., Elderson, M.F., Eriksson, J.G., Evans, D.M., Faul, J.D., Ferrucci, L., Garcia, M.E., Gronberg, H., Guethnason, V., Hall, P., Harris, J.M., Harris, T.B., Hastie, N.D., Heath, A.C., Hernandez, D.G., Hoffmann, W., Hofman, A., Holle, R., Holliday, E.G., Hottenga, J.J., Iacono, W.G., Illig, T., Jarvelin, M.R., Kahonen, M., Kaprio, J., Kirkpatrick, R.M., Kowgier, M., Latvala, A., Launer, L.J., Lawlor, D.A., Lehtimaki, T., Li, J., Lichtenstein, P., Lichtner, P., Liewald, D.C., Madden, P.A., Magnusson, P.K., Makinen, T.E., Masala, M., McGue, M., Metspalu, A., Mielck, A., Miller, M.B., Montgomery, G.W., Mukherjee, S., Nyholt, D.R., Oostra, B.A., Palmer, L.J., Palotie, A., Penninx, B.W., Perola, M., Peyser, P.A., Preisig, M., Raikkonen, K., Raitakari, O.T., Realo, A., Ring, S.M., Ripatti, S., Rivadeneira, F., Rudan, I., Rustichini, A., Salomaa, V., Sarin, A.P., Schlessinger, D., Scott, R.J., Snieder, H., St, P.B., Starr, J.M., Sul, J.H., Surakka, I., Svento, R., Teumer, A., Tiemeier, H., van Rooij, F.J., Van Wagoner, D.R., Vartiainen, E., Viikari, J., Vollenweider, P., Vonk, J.M., Waeber, G., Weir, D.R., Wichmann, H.E., Widen, E., Willemsen, G., Wilson, J.F., Wright, A.F., Conley, D., Davey-Smith, G., Franke, L., Groenen, P.J., Hofman, A., Johannesson, M., Kardia, S.L., Krueger, R.F., Laibson, D., Martin, N.G., Meyer, M.N., Posthuma, D., Thurik, A.R., Timpson, N.J., Uitterlinden, A.G., van Duijn, C.M., Visscher, P.M., Benjamin, D.J., Cesarini, D., & Koellinger, P.D. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science, 340 (6139)*, p. 1467-1471.

Robbers, S.C., van Oort, F.V., Polderman, T.J., Bartels, M., Boomsma, D.I., Verhulst, F.C., Lubke, G.H., & Huizink, A.C. (2011). Trajectories of CBCL Attention Problems in childhood. *European Child and Adolescent Psychiatry, 20 (8)*, p. 419-427.

Rowe, D.C., & Britt, C.L. (1991). Developmental explanations of delinquent behavior among siblings: common factor vs. transmission mechanisms. *Journal of Quantitative Criminology, 7 (4)*, p. 315-332.

Scarr, S., & Weinberg, R.A. (1983). The Minnesota adoption studies: genetic differences and malleability. *Child Development, 54 (2),* p. 260-267.

Scarr-Salapatek, S. (1971). Race, social class, and IQ. *Science, 174 (4016)*, p. 1285-1295.

Schachter, H.M., King, J., Langford, S., & Moher, D. (2001). How efficacious and safe is short-acting methylphenidate for the treatment of attention-deficit disorder in children and adolescents? A meta-analysis. *Canadian Medical Association Journal, 165 (11)*, p. 1475-1488.

Somersalo, H., Solantaus, T., & Almqvist, F. (2002). Classroom climate and the mental health of primary school children. *Nordic Journal of Psychiatry, 56 (4)*, p. 285-290.

Sroufe, L.A. (1990). Considering normal and abnormal together: the essence of developmental psychopathology. *Development and Psychopathology, 2 (4)*, p. 335-347.

Steele, C.M. (1997). A threat in the air: how stereotypes shape intellectual identity and performance. *American Psychologist, 52 (6)*, p. 613-629.

Turkheimer, E., Haley, A., Waldron, M., d'Onofrio, B., & Gottesman, I.I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science, 14 (6)*, p. 623-628.

Turkheimer, E., & Waldron, M. (2000). Nonshared environment: a theoretical, methodological, and quantitative review. *Psychological Bulletin, 126 (1)*, p. 78-108.

van IJzendoorn, M.H., Juffer, F., & Poelhuis, C.W.K. (2005). Adoption and cognitive development: a meta-analytic comparison of adopted and nonadopted children's IQ and school performance. *Psychological Bulletin, 131 (2)*, p. 301-316.

van Langen, A., & Driessen, G. (2006). Sekseverschillen in onderwijsloopbanen: een internationaal comparatieve trendstudie. Nijmegen, The Netherlands: ITS.

van Soelen, I.L., Brouwer, R.M., van Leeuwen, M., Kahn, R.S., Hulshoff Pol, H.E., & Boomsma, D.I. (2011). Heritability of verbal and performance intelligence in a pediatric longitudinal sample. *Twin Research and Human Genetics, 14 (2)*, p. 119-128.

Vinkhuyzen, A.A.E., Van Der Sluis, S., De Geus, E.J.C., Boomsma, D.I., & Posthuma, D. (2010). Genetic influences on 'environmental' factors. *Genes, Brain and Behavior, 9 (3)*, p. 276-287.

Vlug, K.F. (1997). Because every pupil counts: the success of the pupil monitoring system in The Netherlands. *Education and Information Technologies, 2 (4)*, p. 287-306.

Wagenaar, K., Huisman, J., Cohen-Kettenis, P.T., & Delemarre-van de Waal HA (2008). An overview of studies on early development, cognition, and psychosocial well-being in children born after in vitro fertilization. *Journal of Developmental and Behavioral Pediatrics, 29 (3)*, p. 219-230.

Webbink, D., Posthuma, D., Boomsma, D.I., de Geus, E.J., & Visscher, P.M. (2008). Do twins have lower cognitive ability than singletons? *Intelligence, 36 (6)*, p. 539-547.

Wobma, E., & Garssen, J. (2014). CBS - Aantal meerlingen neemt af. CBS Webmagazine 18 mei 2011.

Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., & Kutalik, Z. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics, 46 (1173)*, p. 1173-1186.

Wray, N.R., & Maier, R. (2014). Genetic basis of complex genetic disease: the contribution of disease heterogeneity to missing heritability. *Current Epidemiology Reports 1 (4)*, p. 220-227.

NEDERLANDSE SAMENVATTING

NEDERLANDSE SAMENVATTING

Kinderen verschillen in hoe goed ze het doen op school. Sommige kinderen maken zich basisvaardigheden met gemak eigen en onthouden zonder problemen wat de leerkracht tijdens de les vertelt, terwijl andere kinderen moeite hebben om bij te blijven met de stof. Zelfs kinderen van dezelfde leeftijd die naar dezelfde school gaan en les krijgen van dezelfde leerkracht verschillen in hun prestaties op school. Slechte schoolprestaties zijn belangrijke voorspellers van tegenvallende schoolprestaties op latere leeftijd, schooluitval, criminaliteit (Moilanen, Shaw, & Maxwell, 2010) en van talloze andere ongunstige uitkomsten op volwassen leeftijd, waaronder een lager inkomen (Julian & Kominski, 2011) en een slechtere gezondheid (Mackenbach et al., 1997).

Meerdere onderzoeken hebben aangetoond dat ADHD (aandachtstekort-hyperactiviteitsstoornis) en ODD (oppositioneel opstandige gedragsstoornis) een negatieve invloed hebben op schoolprestaties (Greene et al., 2002; Polderman et al., 2010). Kinderen met ADHD hebben problemen met aandacht vasthouden, hyperactiviteit en/of impulsiviteit, interfereert met het dagelijks functioneren. Wereldwijd komt ADHD voor bij ongeveer 3 tot 7 procent van alle kinderen in de schoolleeftijd (American Psychiatric Association, 2000). Op school hebben deze kinderen bijvoorbeeld moeite om op hun stoel te blijven zitten en om hun aandacht bij de les te houden. ODD wordt gekenmerkt door vijandig en opstandig gedrag, ten opzichte van mensen met gezag, meer dan wat als normaal kindergedrag wordt gezien. Schattingen van de prevalentie van ODD bij kinderen lopen uiteen van 2 tot 16 procent (American Psychiatric Association, 2000). Op school maken deze kinderen vaak ruzie met hun juf of meester en hebben ze vaak woedeaanvallen. Kinderen met deze diagnoses blijven vaker zitten en worden vaker naar het speciaal basisonderwijs verwezen.

Het hoofddoel van dit proefschrift is om de kennis te vergroten over de oorzaken van verschillen tussen basisschoolkinderen in schoolprestaties en van verschillen in ODD en ADHD gerelateerd gedrag. Zowel erfelijke aanleg als de omgeving dragen bij aan de verschillen tussen kinderen in schoolprestaties en in gedrag. Het is bijna nooit óf de genetische aanleg óf de omgeving die bepaalt hoe kinderen zich ontwikkelen, maar een complex samenspel tussen 'nature' (genen) én 'nurture' (omgeving). Er zijn maar weinig eigenschappen waarbij genetische aanleg geen enkele rol speelt. De omgeving waarin kinderen zich bevinden heeft mede invloed op hun ontwikkeling en bepaalt soms ook in hoeverre genetische aanleg de ontwikkeling kan beïnvloeden. Er kan een onderscheid gemaakt worden tussen de gedeelde en de unieke omgeving. Bij gedeelde omgeving moet worden gedacht aan alle invloeden die kinderen uit een gezin meer op elkaar doen lijken dan kinderen die opgroeien in verschillende gezinnen. Unieke omgevingsinvloeden zijn gedefinieerd als invloeden die kinderen van elkaar doen verschillen ook al groeien ze op in hetzelfde gezin en delen ze hun genetisch materiaal.

Tweelingonderzoek geeft informatie over de relatieve invloed van de genen en de omgeving door het feit dat er twee soorten tweelingen bestaan. Eeneiige of mono-zygote (MZ) tweelingen worden geboren als een bevruchte eicel zich, om nog steeds onbekende redenen, binnen een paar dagen na de bevruchting in tweeën splitst. Als gevolg van deze splitsing zijn eeneiige tweelingen genetisch (vrijwel) identiek. Ze zijn dus ook altijd van hetzelfde geslacht. Twee-eiige of di-zygote (DZ) tweelingen ontstaan wanneer na een dubbele ovulatie beide eicellen bevrucht worden. Deze tweelingen zijn genetisch gezien net zo verwant als gewone broers en zussen en delen gemiddeld de helft van hun genetisch materiaal. Maar wat hen anders maakt dan gewone broers en zussen is dat ze onder dezelfde omstandigheden geboren zijn. Ze delen net als eeneiige tweelingen prenatale omstandigheden, hebben dezelfde leeftijd, en groeien op in dezelfde omgeving met dezelfde ouders en in hetzelfde sociaal economische milieu. Daarom vormen twee-eiige tweelingen de perfecte vergelijkingsgroep voor eeneiige tweelingen. Over het algemeen geldt dat hoe groter het verschil tussen de overeenkomsten tussen eeneiige en twee-eiige tweelingen, hoe groter de invloed van genen is. Wanneer eeneiige tweelingen net zoveel op elkaar lijken als twee-eiige tweelingen heeft vooral de gedeelde omgeving invloed. De unieke omgeving is verantwoordelijk voor de verschillen tussen eeneiige tweelingen (Plomin et al., 2008).

De studies in dit proefschrift maken gebruik van de gegevens die in de afgelopen 25 jaar, bij het Nederlands Tweelingen Register (NTR) werden verzameld met hulp van ouders en leerkrachten van 7, 9 en 12-jaar oude twee- en drielingen en de broers en zussen van deze meerlingen. De ouders en de leerkrachten van deze kinderen hebben lijsten ingevuld met vragen over het gedrag en over de schoolprestaties van de kinderen. Daarnaast hebben grote groepen ouders en leerkrachten de resultaten van de Cito leerlingvolgsysteemtoetsen (Cito, 2014; Vlug, 1997) en de Cito eindtoets (Cito, 2002) doorgegeven. Ten slotte heeft een deel van deze gezinnen lichaamsmateriaal afgestaan waar DNA uit geïsoleerd is voor genetisch onderzoek.

In het eerste deel van dit proefschrift worden verschillende voorspellers van schoolprestaties en gedrag onderzocht. In **hoofdstuk 2** wordt gekeken naar de invloed van pre- en perinatale risicofactoren, die vaker voorkomen bij tweelingen dan bij eenlingen, op schoolprestaties. Daarnaast wordt gekeken naar zygositeit, een risicofactor die uniek is voor tweelingen. Tweelingen worden gemiddeld vroeger geboren (Gielen et al., 2010) en hebben gemiddeld een lager geboortegewicht dan eenlingen. Beide factoren zijn eerder al in verband gebracht met de cognitieve ontwikkeling van kinderen. Daarnaast is de vraag of schoolprestaties bij tweelingen afwijken van kinderen die geen tweeling zijn. De tweelingen worden vergeleken met hun eigen broertjes en zusjes, zodat de meeste demografische kenmerken gelijk zijn tussen de groepen. Een laag geboortegewicht en een relatief laag geboortegewicht voor de duur van de zwangerschap hebben een negatieve invloed op de schoolprestaties. Wanneer gecorrigeerd wordt voor het sociaal economische milieu blijft deze samenhang bestaan. De effecten van deze risicofactoren zijn echter relatief klein, zeker wanneer deze worden vergeleken met verschillen tussen jongens en meisjes. De risicofactoren vruchtbaarheidsbehandeling, in de couveuse gelegen en complicaties bij de geboorte hebben alleen invloed op de prestaties bij schoolgym. De andere risicofactoren, bevallingswijze en zygositeit, hebben geen invloed op de schoolprestaties van de tweelingen. De vergelijking van de schoolprestaties van de tweelingen met die van hun broers en zussen, laat kleine verschillen zien op alle inhoudelijke schoolvakken. Deze verschillen tussen tweelingen en eenlingen treden alleen op wanneer de tweelingen met hun oudere broers en zussen worden vergeleken en niet wanneer gekeken wordt naar hun jongere broers en zussen. Dit suggereert dat de geboortevolgorde in een gezin een deel van de gevonden verschillen verklaart tussen tweelingen en eenlingen.

Een andere factor die schoolprestaties zou kunnen beïnvloeden is het geslacht van de leerkracht. Jongens en meisjes kunnen een leerkracht hebben van het andere of van hetzelfde geslacht. Er wordt wel gezegd dat het voor meisjes beter is om een juf te hebben en dat voor jongens een meester beter is. In **hoofdstuk 3** wordt geprobeerd om een empirische bijdrage te leveren aan de discussie of jongens mogelijk benadeeld worden door de feminisering van het basisonderwijs. Hiervoor is gebruik gemaakt van twee bijzondere groepen tweelingen. De eerste groep bestaat uit eeneiige tweelingen van wie het ene kind les heeft van een meester en het andere kind van een juf. De tweede groep bestaat uit twee-eiige tweelingparen van verschillend geslacht, die les hebben van of een juf of een meester. De verschillen binnen een tweelingpaar kunnen worden toegeschreven aan de invloed van het geslacht van de leerkracht, omdat (een deel) van hun genetische aanleg, familieachtergrond, sociaal economische milieu en vele andere kenmerken voor een groot deel gelijk zijn. Hierdoor wordt voor de invloed van deze kenmerken gecontroleerd. Het blijkt dat jongens beter presteren op het gebied van rekenen, terwijl meisjes hoger scoren op het gebied van taal en lezen. Jongens laten ook meer aandachtsproblemen en hyperactiviteit zien in de klas. Deze resultaten staan echter los van het geslacht van de leerkracht.

Om meer inzicht te krijgen in de oorzaken van verschillen tussen kinderen in schoolprestaties en in gedrag wordt in het tweede deel van dit proefschrift gekeken naar de invloed van genetische en omgevingsfactoren op schoolprestaties en de interactie tussen genen en omgeving. In **hoofdstuk 4** wordt de erfelijkheid van de resultaten van de Citotoetsen (groep 3 t/m 8) en van de Cito eindtoets (groep 8) onderzocht. Genetische aanleg heeft meer invloed dan de omgeving op de resultaten voor rekenen (60-74%), lezen (72-82%), begrijpend lezen (54-63%) en spelling (33-70%). De relatieve bijdrage van de genen en de omgeving blijft nagenoeg gelijk gedurende de basisschoolperiode. Een uitzondering hierop is spelling, waar invloed van de erfelijkheid aan het begin van de basisschool een stuk lager is dan in hogere groepen. De invloed van genen op de score op de Cito eindtoets is met 74 procent aanzienlijk; dit betekent dat genetische aanleg voor het overgrote deel de verschillen verklaart tussen kinderen wat betreft hun score op deze toets. In sommige vakken halen jongens betere resultaten, terwijl meisjes het in andere vakken weer beter doen. Meisjes zijn beter in begrijpend lezen en jongens in rekenen. Op de Cito eindtoets scoren jongens beter op de onderdelen rekenen, studievaardigheden en wereldoriëntatie, terwijl meisjes hoger scoren op het onderdeel taal. Ditzelfde patroon van verschillen tussen jongens en meisjes rapporteert het Cito op haar website. De relatieve invloed van de genen en de omgeving is hetzelfde voor jongens en meisjes en ook komen dezelfde genen tot expressie bij jongens en meisjes. Dat de onderzochte verschillen in schoolprestaties voor een groot deel zijn toe te schrijven aan erfelijkheid betekent dat het onderwijssysteem in Nederland relatief homogeen is. Wat niet mag worden vergeten, is dat erfelijkheid niet hetzelfde is als determinisme. De verschillen in schoolprestaties tussen kinderen zijn voor een groot deel erfelijk bepaald, maar het gemiddelde van een groep kinderen kan positief worden beïnvloed door een goede leerkracht en school. De uitdaging voor een leerkracht is om de lessen aan te passen aan de genetische aanleg van kinderen zodat kinderen die moeite hebben op school de basisvaardigheden leren en kinderen die makkelijk leren voldoende worden uitgedaagd.

In **hoofdstuk 5** wordt de erfelijkheid van schoolprestaties in Nederland vergeleken met de bestaande literatuur over tweelingstudies naar schoolprestaties van basisschoolkinderen uit andere landen. We vinden in totaal 61 onderzoeken, die zijn uitgevoerd in 11 verschillende cohorten uit 6, meestal Engelstalige, landen. Om te komen tot de beste schatting van de erfelijkheid, de relatieve bijdrage van de genen aan verschillen tussen kinderen op schoolprestaties, wordt uit alle studies de grootst mogelijke steekproef samengesteld. Aangezien de studies in deze review gebaseerd zijn op data van slechts 11 cohorten komen veel kinderen in meer studies voor. Hierdoor kunnen niet alle studies meegenomen worden in de meta-analyses en wordt gekozen voor de grootste steekproef. Nadat de tweelinggegevens van alle onafhankelijke studies bij elkaar gevoegd waren zijn, wordt met een meta-analyse een gemiddelde erfelijkheid geschat. De resultaten van deze meta-analyse laten een grote erfelijkheid zien voor lezen (73%), begrijpend lezen (49%), rekenen (57%), spelling (44%) en algemene schoolprestaties (66%). Als tweede stap is een heterogeniteitstest gedaan om de verschillen in schattingen van de erfelijkheid tussen de verschillende landen te toetsen. De invloed van de genetische aanleg van een kind verschilt tussen de VS, het Verenigd Koninkrijk (VK) en Nederland. De erfelijkheid voor lezen, begrijpend lezen en rekenen is consistent hoog in Nederland terwijl dit niet het geval is in de VS en het VK. Het land, waarin het kind onderwijs volgt, blijkt een omgevingsfactor te zijn, die het effect van de genen op schoolprestaties beïnvloedt. Dit suggereert dat de schoolomgeving en het onderwijs in Nederland veel gelijkmatiger zijn dan in de VS of het VK.

In **hoofdstuk 6** wordt de erfelijkheid van ODD- en ADHD-gedrag in kaart gebracht voor de leeftijden 7, 9 en 12 jaar. Daarnaast wordt bekeken of de klasomgeving, het geslacht van de leerling en het geslacht van de leerkracht een effect hebben op de erfelijkheid van ODD- en ADHD-gedrag. Dit fenomeen wordt aangeduid als 'genotype-omgeving interactie' en beschrijft dat de mate van erfelijkheid kan afhangen van de omgeving, zoals de klas, waarin het gedrag wordt geobserveerd. Eerst wordt onderzocht of het meetinstrument waarmee ODD- en ADHD-gedrag worden beoordeeld, de Conners' Leerkracht Beoordelingsschaal (Conners, 2001), dezelfde onderliggende gedragsproblemen meet bij jongens en meisjes en gemeten door mannelijke en vrouwelijke beoordelaars. Met andere woorden, de vragen van het meetinstrument moeten meetinvariant zijn en dus hetzelfde meten in alle groepen die onderzocht worden. Wanneer jongens gemiddeld hoger scoren op een bepaalde vraag dan meisjes, terwijl ze niet hoger scoren op het onderliggende gedragsprobleem, dan is het meesinstrument niet meetinvariant. Dit heeft als gevolg dat een zinvolle interpretatie van verschillen tussen jongens en meisjes met een dergelijk meetinstrument veel moeilijker, zo niet onmogelijk, is. Drie van de vier schalen (oppositioneel gedrag, hyperactiviteit en de ADHD index) van de Conners' zijn meetinvariant over 4 groepen; jongens bij een meester, jongens bij een juf, meisjes bij een meester en meisjes bij een juf. Deze schalen kunnen dus gebruikt worden om deze groepen te vergelijken. De schaal 'aandacht/cognitieve problemen' blijkt echter geen homogeen construct te meten in deze vier groepen en dit zet vraagtekens bij de betrouwbaarheid en het gebruik in de praktijk van deze schaal. In dit proefschrift is deze schaal daarom verder niet onderzocht. De erfelijkheid van ODD- en ADHD-gedrag blijkt groter te zijn in bij tweelingen die in dezelfde klas zitten in vergelijking met tweelingen die in verschillende klassen zitten. Dit laat zien dat verschillen in gedrag deels bepaald worden door verschillende klasomgevingen, leerkrachten en leeftijdsgenootjes. In sommige gevallen heeft het geslacht van de leerkracht invloed op de mate waarin de individuele verschillen in ODD- en ADHD-gedrag worden verklaard door genen of door de omgeving. Echter zijn deze verschillen niet consistent voor de verschillende schalen en leeftijden wat interpretatie moeilijk maakt. De relatieve invloed van de genen en de omgeving verschilt tussen jongens en meisjes voor hyperactiviteit en oppositioneel opstandig gedrag, maar niet voor ADHD-gedrag.

In het derde deel van dit proefschrift wordt de samenhang van gedragsproblemen met schoolprestaties onderzocht door te kijken of dezelfde genetische varianten een rol spelen bij schoolprestaties en gedragsproblemen. Kinderen die meer ODD- of ADHD-gedrag laten zien, presteren over het algemeen slechter op school. Een belangrijke vraag is wat de oorzaak van deze samenhang tussen gedrag en schoolprestaties is. In **hoofdstuk 7** wordt onderzocht of genetische varianten die worden geassocieerd met het opleidingsniveau bij volwassenen ook geassocieerd worden met schoolprestaties en gedragsproblemen bij kinderen. Hiervoor wordt een polygenetische risicoscore gedefinieerd op grond van het genotype van het kind en alle effecten van de genetische varianten zoals gevonden in de studie onder volwassenen. Het blijkt dat deze genetische aanleg voorspellend is voor de schoolprestaties (4% verklaarde variantie) van kinderen en voor de mate van ADHD-gedrag thuis en op school (2% verklaarde variantie). Dit laat op DNA niveau zien dat er genetische varianten zijn die een invloed hebben op zowel de prestaties van een kind op school als op ADHD-gedrag.

Een genetische overlap tussen ADHD-gedrag en schoolprestaties, zoals gevonden wordt in dit proefschrift, kan worden veroorzaakt door genetische pleiotropy, of door een causaal verband. Genetische pleiotropy beschrijft het fenomeen dat dezelfde genetische varianten een invloed hebben op meerdere eigenschappen. Bij een causaal effect zullen de genetische varianten met een invloed op ADHD ook indirect lagere schoolprestaties tot gevolg hebben. Een kind met een genetische aanleg voor ADHD zal zich hierdoor moeilijker kunnen concentreren tijdens de les, wat uiteindelijk leidt tot slechtere schoolprestaties. In **hoofdstuk 8** wordt getoetst of een causaal model of genetische pleiotropy de beste verklaring is voor de samenhang tussen ODD gedrag en schoolprestaties en tussen ADHD-gedrag en schoolprestaties. Als 12-jarige kinderen, volgens hun moeder, meer ODD- of ADHD-gedrag vertonen, scoren ze ook lager op de Cito eindtoets. Ook probleemgedrag gemeten op 7-jarige leeftijd blijkt voorspellend te zijn voor een lagere score op de Cito eindtoets. De samenhang tussen ODD gedrag en schoolprestaties is zwak waardoor de resultaten door een gebrek aan statistische power niet consistent zijn. Als gevolg hiervan kan een causaal effect niet worden bevestigd noch ontkracht. De resultaten wat betreft het ADHD-gedrag zijn veel duidelijker. Dat een causaal effect de meest waarschijnlijke verklaring is voor de samenhang tussen schoolprestaties en ADHD-gedrag, blijkt uit de bevindingen in een groep van genetisch identieke (eeneiige) tweelingen. In de groep eeneiige tweelingen presteert het kind dat het meeste ADHD-gedrag laat zien, ook slechter op school dan zijn of haar, genetisch identieke, tweelingbroer of zus. Ook de bevinding dat alle genetische factoren en alle omgevingsfactoren die ADHD-gedrag beïnvloeden ook invloed hebben op de schoolprestaties, pleit voor een causaal effect voor de samenhang tussen ADHD en schoolprestaties. Dit betekent dat wanneer een gedragsinterventie of het voorschrijven van medicijnen het ADHD-gedrag vermindert, de schoolprestaties van het kind ook zouden moeten verbeteren.

REFERENTIES

American Psychiatric Association (2000). Diagnostic and Statistical Manual of Mental Disorders: 4th ed., text rev. DSM-IV-TR. Washington, DC: APA.

Bronfenbrenner, U., & Ceci, S.J. (1994). Nature-nurture reconceptualized in developmental perspective: a bioecological model. *Psychological Review, 101 (4),* p. 568-586.

Cito (2002). Eindtoets Basisonderwijs. Arnhem, The Netherlands: Cito.

Cito (2014). Cito Leerling Volgsysteem. Arnhem, The Netherlands: Cito.

Conners, C.K. (2001). Conners' rating scales - revised. New York, NY: Multi-Health Systems, Inc.

Gielen, M., van Beijsterveldt, C.E., Derom, C., Vlietinck, R., Nijhuis, J.G., Zeegers, M.P., & Boomsma, D.I. (2010). Secular trends in gestational age and birthweight in twins. *Human Reproduction, 25 (9),* p. 2346-2353.

Greene, R.W., Biederman, J., Zerwas, S., Monuteaux, M.C., Goring, J.C., & Faraone, S.V. (2002). Psychiatric comorbidity, family dysfunction, and social impairment in referred youth with oppositional defiant disorder. *American Journal of Psychiatry, 159 (7),* p. 1214-1224.

Julian, T., & Kominski, R. (2011). Education and synthetic work-life earnings estimates. American Community Survey Report: Census.

Mackenbach, J.P., Kunst, A.E., Cavelaars, A.E., Groenhof, F., & Geurts, J.J. (1997). Socioeconomic inequalities in morbidity and mortality in western Europe. The EU Working Group on Socioeconomic Inequalities in Health. *Lancet, 349 (9066),* p. 1655-1659.

Moilanen, K.L., Shaw, D.S., & Maxwell, K.L. (2010). Developmental cascades: externalizing, internalizing, and academic competence from middle childhood to early adolescence. *Development and Psychopathology, 22 (3),* p. 635-653.

Plomin R., DeFries J.C., McClearn G.E., & McGuffin P.s (2008). Behavioral genetics.*5th Edition.* New York, NY: Worth Publishers.

Polderman, T.J.C., Boomsma, D.I., Bartels, M., Verhulst, F.C., & Huizink, A.C. (2010). A systematic review of prospective studies on attention problems and academic achievement. *Acta Psychiatrica Scandinavica, 122 (4),* p. 271-284.

Rende, R., & Plomin, R. (1992). Diathesis-stress models of psychopathology: A quantitative genetic perspective. *Applied and Preventive Psychology, 1 (4),* p. 177-182.Vlug, K.F. (1997). Because every pupil counts: the success of the pupil monitoring system in The Netherlands. *Education and Information Technologies, 2 (4),* p. 287-306.

APPENDICES

APPENDIX A DATA COLLECTION

The Netherlands Twin Register (NTR) was established around 1987 by the Department of Biological Psychology at the VU University Amsterdam with the main goal to investigate differences in general cognitive ability, psychopathology and physical and psychological well-being between people. The NTR is divided into two parts: the Young Netherlands Twin Register (Y-NTR), twins from birth until the children reach the age of 18 years and the Adult Netherlands Twin Register (A-NTR), adult twins and their family members and spouses. This thesis is primarily based on data provided by the primary school teachers of young twins and their non-twin siblings. The twins from the Y-NTR are mainly recruited via a commercial organization that provides gift boxes for parents of newborns and through the Dutch association for parents of multiples (NVOM). The Y-NTR receives the address information from families with a multiple birth and contacts them with a request for registration by sending the parents a registration form, the yearly news bulletin of the NTR (TWINfo) and the first survey with questions about the pregnancy, delivery and early development of the children.

Parents and teachers are the main informants for young children, whereas at later ages data are collected through self-reports. Data collection in the Y-NTR is based on the birth cohort of the multiples. Parents of the multiples receive a survey about the development of their children when their twins are 1, 2, 3, 5, 7, 9 and 12 years. At the ages 7, 9 and 12 years, when the children are attending primary school, parents are asked consent to approach the teacher(s) of their children with a survey. During adolescence, at the ages 14, 16 and (previously) 18 years, the children and their siblings are invited to complete a self-report survey, after permission is obtained from the parents. When the children that are part of the Y-NTR reach the age of 18 years, they are invited to take part in the research of the A-NTR.

Since 1999 parents of twins and since 2009 parents of triplets are asked for their consent to approach the teacher(s) of their multiples) with a survey. From 2007 onwards, parents of twins are asked for consent to send a survey to the teacher of the non-twin siblings (6-12 years) of the twin. The data collection of the teachers is ongoing and each year, the parents of children that are approximately 7, 9 and 12 years old are contacted to ask for their permission. In November, parents are approached for their consent by (e-)mail and asked to provide the contact details of the teachers. This information is entered into the personal information and addresses database (Personen en Adresseninformatiesysteem Nederlands Tweelingen Register (Panter)) and teachers are invited by mail to participate. The survey for the primary school teachers includes items on background information of the teacher, class and school, functioning of the child at school, school performance, bullying and standardized questionnaires, namely the Teacher Report Form (TRF) (Achenbach, 1991) (see Appendix B) and the short version of the Conners' Teacher Ratings Scale - Revised (CTRS-R) (Conners et al., 1998; Conners, 2001) (see Appendix C). Over the years, different versions of the teacher survey have been collected. The short CTRS-R was not included in the first version of the teacher survey. An update of the TRF with several new items was published in 2006 and added to a new version of the survey for teachers. Some questions and/or their answer categories have been changed, were removed or have been added through the years. All returned surveys are collected in harmonized databases organized by age at the time of data collection. These databases from different ages can be linked through anonymous identification numbers.

Initially, from 2000 onwards, results on a standardized educational achievement test, which is administered in the last grade of primary school, were also obtained from the teachers (Cito, 2002). Later, because results only become available near the end of the school years, parents were asked to report the scores of their children on this test (see Appendix D). In addition, since 2008, teachers are requested to provide information on results of tests that are developed to measure the educational development of children throughout primary school (pupil monitoring system) (Cito, 2014b) (see Appendix E).

PROCEDURES

PARENTAL CONSENT

Before inviting the primary school teachers of twins (and triplets) and their siblings to participate in the study, parents are contacted by (e-)mail to ask for permission. Parents are invited to give their consent at each wave of the data collection regardless of their previous participation. Initially, collection of the parental consent was done by mail, but for the last 5 school years, the consents have been gathered via online forms. Parents are contacted with a letter through regular mail or via e-mail (see Appendix F) in which they are asked to fill out the online permission form and indicate whether or not they give permission for their twin (or triplet) and/or siblings. Upon request, a paper-and-pencil version of the parental consent form is sent to the parents who do not want to or are unable to fill out the online form (see Appendix G). If parents give consent, they are requested to provide the names and contact details of the teachers of the children. When the parents do not know the surname of the teacher, the school is contacted to make sure the teachers are addressed correctly. When parents do not fill out the online parental consent within 6-8 weeks, they are contacted again by regular mail with a reminder. The procedure to contact the parents for permission remained relatively unchanged over the recent years with the exception of last year when parents were contacted by phone for a second reminder. The response rate for the parental consent, irrespective of whether parents gave permission to approach the teachers of their children or not, was 37% for the school year 2010-2011, 43% for 2011-2012, 42% for 2012-2013 and 59% for 2013-2014.

TEACHER SURVEY

From 2012 onwards, the teacher survey is collected online and teachers are invited to participate with a letter send by regular mail containing the hyperlink to the online survey and a personal long-in code and password (see Appendix H). A TWINfo is sent along with the invitation to provide some background information on NTR studies. Teachers are referred to the website for more details about the research conducted with data collected via the teacher survey (see appendix I). One or more teachers, depending on whether the twins go to the same class/school and/or have siblings, are approached to fill out the survey. If a teacher prefers a paper-and-pencil version of the survey, he or she can contact us. When the survey is not completed within 6-8 weeks, teachers receive a reminder letter with the hyperlink and personal codes as well as a paper-and-pencil version of the survey. To compare the willingness to respond to the online teacher survey compared to the paper-and-pencil version, the cohort of 9-year-olds was randomly split into half in the first year that the online survey became available,. One half was invited to fill out the survey online and was reminded with a paper-and-pencil version and a letter containing the hyperlink to the online version of the survey. The other half was invited to fill out the paper-and-pencil version and reminded with a letter with a hyperlink. The initial response rate (before the reminder) was approximately equal for the paper-and-pencil version and the online version (van Beijsterveldt et al., 2013). The response rate was 64% for the school year 2010-2011, 61% for 2011-2012, 54% for 2012-2013 and 63% for 2013-2014. In recent years, more effort is needed to reach the same response rates as teachers have increasingly more work to do alongside teaching.

PUPIL MONITORING SYSTEM

In the Dutch primary school education system, tests are available for all important educational domains that are independent of teaching methods and they can be used to monitor a child's educational development at school. (Cito, 2014b). The tests are administered at fixed time points (i.e. beginning, half way and end of a school year) in each grade and entered into a database which enables teachers to compile student reports with the results of the different tests across all grades. Norms and questions in the tests are updated regularly resulting in multiple versions of the same test. To account for differences between the older and newer versions of the pupil monitoring tests, during data entry, it is thoroughly registered which version of the test is administered. Before starting the data collection of pupil monitoring test results, teachers were asked to indicate which tests provide the best indication of a child's educational achievement. The inventory revealed that the tests measuring arithmetic, reading, reading comprehension and spelling were the most informative (Polderman et al., 2011). Teachers are therefore asked to include an overview of the results of the children on those tests administered halfway throughout each grade. Approximately one third of the teachers who return the teacher survey also send in a student report with the pupil monitoring test results (see Appendix E). Unfortunately, because parents have to give permission to approach the teachers of their children and teachers have to fill out the teacher survey and send in the student reports, the response rate for the pupil monitoring system depends on the response rate of the parental consent, the response rate of the teacher survey and the response rate of the student reports.

SAMPLE

Table 1 shows the total number of surveys returned by the teachers (N = 33.334) at the ages 7, 9 or 12 years. Each survey concerns one child of a twin pair (or triplet) or a non-twin sibling of a twin pair, born between 1986 and 2006. One teacher survey was available for 66% of the children, there were two surveys available for 27% of the children and for 7% of the children, all three teacher surveys were available. The studies presented in this thesis are based on different subsamples of the data as the data collection is an ongoing process and the sample has become larger during the past years.

TABLE 1 The number of teacher surveys (TRF) returned per birth cohort and age of assessment

	Age 7			Age 9			Age 12		
	Twins	**Siblings**	**Triplets**	**Twins**	**Siblings**	**Triplets**	**Twins**	**Siblings**	**Triplets**
1986							82		
1987							596		
1988							540		
1989				432			490		
1990				991			597		
1991				870			612		
1992	592			929			565		
1993	1227	2		959			540		
1994	1168	10		813	24		486	1	
1995	1213	32		720	70		745	2	
1996	1167	74		745	68		567	7	6
1997	993	117		980	63		552	22	21
1998	823	134		763	101	14	510	37	19
1999	979	80		279	99	20	626	55	30
2000	1006	110	26	847	133	11	449	58	13
2001	450	139	35	660	111	13		60	
2002	205	110	18	663	56	12		52	
2003	851	50	9	499	32	20		30	
2004	671	7	15		32			10	
2005	556	2	18		24			8	
2006					4				
Total	**11901**	**867**	**121**	**11150**	**817**	**90**	**7957**	**342**	**89**

MEASUREMENTS

BEHAVIORAL AND EMOTIONAL PROBLEMS

The Teacher Report Form (TRF) consists of 120 items scored on a 3 point scale from 0 (not at all) to 2 (very true or often), measuring behavioral and emotional problems. Teachers have to indicate whether or not a child displayed a certain type of behavior 1) currently or 2) in the last two months. The TRF includes 8 small band syndromes (Rule Breaking Behavior, Aggression, Attention Problems, Withdrawal, Anxiety/Depression, Somatic Complaints, Social Problems and Thought Problems) and 2 broad band scales describing Internalizing Problems and Externalizing Problems. The short Conners' Teacher Rating Scale - Revised (CTRS-R) consists of 28 items scored on a 4 point scale from 0 (not true or never) to 3 (completely true or very often), measuring Oppositional Defiant Disorder (ODD) and Attention Deficit Hyperactivity Disorder (ADHD) behavior. Teachers have to indicate whether a child displayed a certain type of behavior 1) currently or 2) in the prior month. The short version of the CTRS-R includes 4 scales measuring Oppositional Behavior, Cognitive Problems/Inattention, Hyperactivity and a Attention Deficit Hyperactivity Disorder Index. Sum scores are computed only when subjects had no or a limited number of missing items on a scale. A missing item on a scale was then imputed by the averaged item score within a scale of an individual child.

SCHOOL PERFORMANCE

Teacher ratings measure the school performance in four educational domains, namely arithmetic, language, reading and physical education, with two versions of the teacher survey. In the first version, teachers could choose up to six educational domains and rate the proficiency of the students on a five-point scale from 1 (insufficient) to 5 ((very) good). In the second version, teachers rated the proficiency of the students in four predefined educational domains on the same five-point scale.

EDUCATIONAL ACHIEVEMENT

The national test of educational achievement is administered in the final grade of primary school at approximately 80 per cent of all schools in the Netherlands (Cito, 2002). This final test measures what a child has learned in 6 years of primary education. In the Dutch educational system, the results of this test are often used, besides the advice of the teacher, to determine the level of secondary education suitable for a child. The test consists of multiple choice items in four different domains, namely Arithmetic, Language, Study Skills and Science and Social Studies. The first three test scales are combined into a Total Score, which is standardized on a scale from 500 and 550. The Arithmetic scale includes items on numbers and operations, ratios, fractions and percentages, and measurements, geometry, time and money. The Language scale includes items on writing, spelling, reading comprehension and vocabulary. The Study Skills scale includes items on handling of study texts, handling of information, reading diagrams, tables and graphs and map reading. The Science and Social Studies scale includes items on geography, history and biology.

The pupil monitoring system includes, amongst others, Arithmetic, Reading, Reading comprehension and Spelling tests (Cito, 2014a; Vlug, 1997). The Arithmetic test (grade 1 to 6) consists of a part in which children have to solve simple math problems within a short time and a part with more complex math problems without a time limit. The test assesses general knowledge of mathematics and arithmetic and comprises written computational problems of addition, subtraction, multiplication and division and problems on the notion of measurements, time and money and knowledge about fractions, ratios and percentages. The Reading test (grade 1 to 6) measures word decoding skills by counting the total number of individual words a child can correctly read aloud in 1 minute. The test consists of three levels of increasing difficulty and complexity. The first level includes words that are pronounced exactly as they are spelled, the second level includes also other monosyllabic words and the third level includes two or more syllabic words. The study in this thesis uses the most difficult level of the test which is almost never administered in the first grade. The Reading Comprehension test (grade 3 to 6) includes a large variety of different text types and genres with two different types of multiple choice questions. The tests consists of a part in which a child has to read a number of short texts and answer questions related to the text and a part with parts of the text left blank that need to be filled out. The questions aim to assess different components of reading processing by questions regarding both the facts and events described in the texts as well as questions about the purpose of the writer and the intended readership of the texts. The Spelling test (grade 1 to 6) measures both active spelling (writing down the words) and passive spelling (recognizing spelling errors). Active spelling is measured with a dictation by the teacher where a sentence is read aloud and a child has to write down a specific word from this sentence. Passive spelling is measured with multiple choice questions where a student has to choose the sentence in which the bolded word is spelled incorrectly.

REFERENCES

Achenbach, T.M. (1991). Manual for the Child Behavior Checklist/4 - 18 and 1991 Profile. Burlington, VT: University of Vermont Department of Psychiatry.

Cito (2002). Eindtoets Basisonderwijs. Arnhem, The Netherlands: Cito.

Cito (2014a). Cito Leerling Volgsysteem. Arnhem, The Netherlands: Cito.

Cito (2014b). Cito Volgsysteem. Arnhem, The Netherlands: Cito.

Conners, C.K. (2001). Conners' rating scales - revised. New York, NY: Multi-Health Systems, Inc.

Conners, C.K., Sitarenios, G., Parker, J.D., & Epstein, J.N. (1998). Revision and restandardization of the Conners Teacher Rating Scale (CTRS-R): factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology*, *26 (4)*, p. 279-291.

Polderman, T.J., Huizink, A.C., Verhulst, F.C., van Beijsterveldt, C.E., Boomsma, D.I., & Bartels, M. (2011). A genetic study on attention problems and academic skills: Results of a longitudinal study in twins. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *20 (1)*, p. 22

van Beijsterveldt, C.E., Groen-Blokhuis, M., Hottenga, J.J., Franic, S., Hudziak, J.J., Lamb, D., Huppertz, C., de Zeeuw, E., Nivard, M., Schutte, N., Swagerman, S., Glasner, T., van Fulpen, M., Brouwer, C., Stroet, T., Nowotny, D., Ehli, E.A., Davies, G.E., Scheet, P., Orlebeke, J.F., Kan, K.J., Smit, D., Dolan, C.V., Middeldorp, C.M., de Geus, E.J., Bartels, M., & Boomsma, D.I. (2013). The Young Netherlands Twin Register (YNTR): Longitudinal Twin and Family Studies in Over 70,000 Children. *Twin Research and Human Genetics*, *16 (1)*, p. 252-267.

Vlug, K.F. (1997). Because every pupil counts: the success of the pupil monitoring system in The Netherlands. *Education and Information Technologies*, *2 (4)*, p. 287-306.

APPENDIX B TEACHER REPORT FORM

Nederlands Tweelingen- en Familieonderzoek

in verband met privacybescherming van de invuller en leerling zal dit voorblad bij binnenkomst op het NTR verwijderd en apart gearchiveerd worden

Nederlands Tweelingen Register

Vragenlijst voor Leerkrachten

Informatie over de invuller

Deze vragenlijst werd ingevuld door:

Uw voornaam ______

Uw achternaam ______

Uw geslacht ☐1 man ☐2 vrouw

Uw geboortedatum ___-___-______

Invuldatum ___-___-______

Uw relatie tot de leerling:

☐1 Leerkracht

☐2 Speciale leerkracht (toelichting): ______

☐3 Klassenassistent

☐4 Anders (toelichting): ______

Als de leerling een deel van een tweelingpaar is, zitten beide kinderen bij u in de klas? ☐1 nee ☐2 ja ☐3 n.v.t.

Heeft u het onderstaande kind eerder in uw klas gehad? ☐1 nee ☐2 ja, in groep ___

Informatie over de leerling

Voornaam leerling ______

Geslacht leerling ☐1 jongen ☐2 meisje

Naam school ______

Postcode school __ __ __ __ (alleen de cijfers)

E-mailadres school ______

Telefoonnummer school __ __ __ __ __ __ __ __ __ __

Informatie over de school

Type school

☐1 openbaar

☐2 evangelisch onderwijs

☐3 gereformeerd onderwijs

☐4 hindoeïstisch onderwijs

☐5 islamitisch onderwijs

☐6 joods onderwijs

☐7 oecumenisch onderwijs

☐8 protestants-christelijk onderwijs

☐9 reformatorisch onderwijs

☐10 rooms-katholiek onderwijs

☐11 anders, nl ______

Type onderwijs

☐1 regulier

☐2 dalton

☐3 jenaplan

☐4 montessori

☐5 freinet

☐6 vrije school

☐7 anders, nl ______

☐8 speciaal basisonderwijs

☐9 speciaal onderwijs

- ☐9a Cluster 1: Voor blinde of slechtziende kinderen
- ☐9b Cluster 2: Voor dove of slechthorende kinderen, kinderen met ernstige spraakmoeilijkheden
- ☐9c Cluster 3: Voor kinderen met lichamelijk en/of verstandelijke beperkingen, ZMLK en langdurig zieke kinderen met een lichamelijke handicap
- ☐9d Cluster 4: Voor kinderen met ernstige psychiatrische en/of gedragsproblemen

GEDRAGSVRAGENLIJST VOOR LEERKRACHTEN

Uw antwoorden zullen gebruikt worden om deze leerling te vergelijken met andere leerlingen waarover leerkrachten dezelfde formulieren hebben ingevuld. De informatie die u geeft op dit formulier zal ook gebruikt worden ter vergelijking met andere informatie over dezelfde leerling. Graag de vragen zo goed als u kunt beantwoorden, ook als u het idee hebt dat u niet volledig op de hoogte bent. De antwoorden op afzonderlijke vragen zullen gecombineerd worden om meer algemene patronen van gedrag te beschrijven. **Graag met blokletters invullen en alle vragen beantwoorden.**

Informatie over de leerling	
Voornaam leerling	....
Geslacht leerling	☐1 jongen ☐2 meisje
Invuldatum	...... -......-..........
Leeftijd leerling	...... jaar
Groep	......

1.	Hoe lang kent u deze leerling?	☐1 minder dan 2 mnd	☐2 2 – 6 mnd	☐3 meer dan 6 mnd
2.	Hoe goed kent u hem/haar?	☐1 niet zo goed	☐2 gemiddeld	☐3 erg goed
3.	Hoeveel tijd per week zit hij/zij bij u in de klas of ziet u deze leerling?	☐1 minder dan 1 dag	☐2 1-3 dagen	☐3 4-5 dagen
4.	Wat voor les, vak of hulp geeft u aan deze leerling?	☐1 alle vakken	☐2 enkele vakken	☐3 1 vak
5.	Is hij/zij ooit blijven zitten?	☐1 weet ik niet	☐2 nee	☐3 ja, in groep ___
6.	Heeft hij/zij wel eens professionele hulp gehad vanwege problemen in de ontwikkeling?	☐1 weet ik niet	☐2 nee	☐3 ja, in welke vorm: ....

7. Kunt u voor de onderstaande schoolvakken het aantal uren per week aangeven en hoe de resultaten van deze leerling over het algemeen zijn?

	Uren per week	Onvoldoende	Bijna Voldoende	Voldoende	Ruim Voldoende	Goed/ Zeer goed
a. Rekenen	....	☐1	☐2	☐3	☐4	☐5
b. Taal	....	☐1	☐2	☐3	☐4	☐5
c. Lezen	....	☐1	☐2	☐3	☐4	☐5
d. Gymnastiek	....	☐1	☐2	☐3	☐4	☐5

8. Vergeleken met de gemiddelde leerling van dezelfde leeftijd

	Veel minder	Wat minder	Klein beetje minder	Gemiddeld	Klein beetje meer	Wat meer	Veel meer
a. Hoe hard werkt hij/zij?	☐1	☐2	☐3	☐4	☐5	☐6	☐7
b. Hoe adequaat gedraagt hij/zij zich?	☐1	☐2	☐3	☐4	☐5	☐6	☐7
c. Hoeveel gaat hij/zij vooruit met leren?	☐1	☐2	☐3	☐4	☐5	☐6	☐7
d. Hoe opgewekt is hij/zij?	☐1	☐2	☐3	☐4	☐5	☐6	☐7
e. Hoe taakgericht is hij/zij?	☐1	☐2	☐3	☐4	☐5	☐6	☐7
f. Hoe atletisch is hij/zij?	☐1	☐2	☐3	☐4	☐5	☐6	☐7

9a. Hoeveel leerlingen telt uw school? ______ leerlingen

9b. Met hoeveel kinderen zit de leerling het grootste deel van de tijd in de klas? ______ kinderen

9c. Als de leerling in een combinatieklas zit, met hoeveel kinderen zit de leerling in zijn/haar groep (leerjaar)? ______ kinderen

10a. Hoe zou u de samenstelling van de groep waarin de leerling zit beschrijven?
- ☐ 1 bijna geheel Nederlandstalig
- ☐ 2 in meerderheid Nederlandstalig
- ☐ 3 ongeveer voor de helft Nederlandstalig
- ☐ 4 in meerderheid anderstalig
- ☐ 5 bijna geheel anderstalig

10b. Hoe zou u de samenstelling van de groep waarin de leerling zit beschrijven?
- ☐ 1 bijna geheel van hoog opgeleide ouders
- ☐ 2 in meerderheid van hoog opgeleide ouders
- ☐ 3 ongeveer even veel van hoog als van laag opgeleide ouders
- ☐ 4 in meerderheid van laag opgeleide ouders
- ☐ 5 bijna geheel van laag opgeleide ouders

11. Heeft uw groep meerdere leerkrachten? ☐ 1 nee ☐ 2 ja

12. Heeft de school een volcontinu rooster (gezamenlijk overblijven)? ☐ 1 nee ☐ 2 ja

13. Blijft de leerling tussen de middag op school? ☐ 1 nee ☐ 2 ja, ______ dagen per week

14a. Heeft de leerling een indicatie voor een "rugzakje"? ☐ 1 nee ☐ 2 ja

14b. Zo ja, voor welke problematiek? (schrijf op) ____
____

15. Hoe zou u het welbevinden van de leerling beschrijven?
- ☐ 1 altijd of bijna altijd gelukkig
- ☐ 2 vaker gelukkig dan ongelukkig
- ☐ 3 even vaak gelukkig als ongelukkig
- ☐ 4 vaker ongelukkig dan gelukkig
- ☐ 5 altijd of bijna altijd ongelukkig

16. Hoe vaak is deze leerling in de afgelopen maanden…

		niet	1 of 2 keer	2 of 3 keer per maand	ongeveer 1 keer per week	meerdere keren per week
a.	gepest (algemeen)	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
b.	gepest door belediging, uitschelden, of uitlachen? (verbaal)	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c.	gepest door spugen, slaan, schoppen of knijpen? (fysiek)	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
d.	gepest door buitensluiten, negeren of roddelen? (relationeel)	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

17. Hoe vaak heeft deze leerling in de afgelopen maanden…

		niet	1 of 2 keer	2 of 3 keer per maand	ongeveer 1 keer per week	meerdere keren per week
a.	andere leerlingen gepest (algemeen)	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
b.	andere leerlingen gepest door belediging, uitschelden, of uitlachen? (verbaal)	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c.	andere leerlingen gepest door spugen, slaan, schoppen of knijpen? (fysiek)	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
d.	andere leerlingen gepest door buitensluiten, negeren of roddelen?(relationeel)	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Graag met blokletters invullen. Graag alle vragen beantwoorden.

Hieronder staat een lijst met vragen over leerlingen. Alle vragen gaan over hoe deze leerling nu is of in **de afgelopen twee maanden** is geweest. Maak het vakje met de 2 zwart als de vraag **duidelijk of vaak** bij deze leerling past. Maak het vakje met de 1 zwart als de vraag **een beetje** of soms bij deze leerling past. Als de vraag **helemaal niet** bij deze leerling past, maak dan het vakje met de 0 zwart. Beantwoord alle vragen zo goed als u kunt, ook al lijken sommige vragen niet bij deze leerling te passen.

0 = Helemaal Niet (voor zover u weet) **1 = Een Beetje of Soms** **2 = Duidelijk of Vaak**

- $\square_0$ $\square_1$ $\square_2$ 1. Doet te jong voor zijn/haar leeftijd
- $\square_0$ $\square_1$ $\square_2$ 2. Bromt of maakt andere vreemde geluiden tijdens de les
- $\square_0$ $\square_1$ $\square_2$ 3. Maakt veel ruzie
- $\square_0$ $\square_1$ $\square_2$ 4. Maakt dingen waar hij/zij mee begint niet af
- $\square_0$ $\square_1$ $\square_2$ 5. Er is heel weinig wat hij/zij leuk vindt
- $\square_0$ $\square_1$ $\square_2$ 6. Uitdagend, geeft brutale antwoorden
- $\square_0$ $\square_1$ $\square_2$ 7. Schept op, doet stoer
- $\square_0$ $\square_1$ $\square_2$ 8. Kan zich niet concentreren, kan niet lang de aandacht ergens bij houden
- $\square_0$ $\square_1$ $\square_2$ 9. Kan bepaalde gedachten niet uit zijn/haar hoofd zetten, obsessies (schrijf op): ………………………………
- $\square_0$ $\square_1$ $\square_2$ 10. Kan niet stilzitten, onrustig of hyperactief
- $\square_0$ $\square_1$ $\square_2$ 11. Klampt zich vast aan volwassenen of is te afhankelijk
- $\square_0$ $\square_1$ $\square_2$ 12. Klaagt over zich eenzaam voelen
- $\square_0$ $\square_1$ $\square_2$ 13. In de war of wazig denken
- $\square_0$ $\square_1$ $\square_2$ 14. Huilt veel
- $\square_0$ $\square_1$ $\square_2$ 15. Friemelt, wiebelt
- $\square_0$ $\square_1$ $\square_2$ 16. Wreed, pesterig of gemeen tegen anderen
- $\square_0$ $\square_1$ $\square_2$ 17. Dagdromen of gaat op in zijn/haar gedachten
- $\square_0$ $\square_1$ $\square_2$ 18. Verwondt zichzelf opzettelijk of doet zelfmoordpogingen
- $\square_0$ $\square_1$ $\square_2$ 19. Eist veel aandacht op
- $\square_0$ $\square_1$ $\square_2$ 20. Vernielt eigen spullen
- $\square_0$ $\square_1$ $\square_2$ 21. Vernielt spullen van anderen
- $\square_0$ $\square_1$ $\square_2$ 22. Vindt het moeilijk om aanwijzingen op te volgen
- $\square_0$ $\square_1$ $\square_2$ 23. Is ongehoorzaam op school
- $\square_0$ $\square_1$ $\square_2$ 24. Stoort andere leerlingen
- $\square_0$ $\square_1$ $\square_2$ 25. Kan niet goed opschieten met andere leerlingen
- $\square_0$ $\square_1$ $\square_2$ 26. Lijkt zich niet schuldig te voelen na zich misdragen te hebben
- $\square_0$ $\square_1$ $\square_2$ 27. Snel jaloers
- $\square_0$ $\square_1$ $\square_2$ 28. Overtreedt schoolregels
- $\square_0$ $\square_1$ $\square_2$ 29. Is bang voor bepaalde dieren, situaties of plaatsen anders dan school (schrijf op): ………………………………
- $\square_0$ $\square_1$ $\square_2$ 30. Is bang om naar school te gaan
- $\square_0$ $\square_1$ $\square_2$ 31. Is bang dat hij/zij iets slechts zou kunnen doen of denken
- $\square_0$ $\square_1$ $\square_2$ 32. Heeft het gevoel dat hij/zij perfect moet zijn
- $\square_0$ $\square_1$ $\square_2$ 33. Heeft het gevoel of klaagt erover dat niemand van hem/haar houdt
- $\square_0$ $\square_1$ $\square_2$ 34. Heeft het gevoel dat anderen hem/haar te pakken willen nemen
- $\square_0$ $\square_1$ $\square_2$ 35. Voelt zich waardeloos of minderwaardig
- $\square_0$ $\square_1$ $\square_2$ 36. Bezeert zich vaak, krijgt vaak ongelukken
- $\square_0$ $\square_1$ $\square_2$ 37. Vecht veel
- $\square_0$ $\square_1$ $\square_2$ 38. Wordt veel gepest
- $\square_0$ $\square_1$ $\square_2$ 39. Gaat om met jongens of meisjes die in moeilijkheden raken
- $\square_0$ $\square_1$ $\square_2$ 40. Hoort geluiden of stemmen die er niet zijn (schrijf op): ………………………………
- $\square_0$ $\square_1$ $\square_2$ 41. Impulsief of doet dingen zonder er bij na te Denken
- $\square_0$ $\square_1$ $\square_2$ 42. Is liever alleen dan met anderen
- $\square_0$ $\square_1$ $\square_2$ 43. Liegt of bedriegt
- $\square_0$ $\square_1$ $\square_2$ 44. Bijt nagels
- $\square_0$ $\square_1$ $\square_2$ 45. Nerveus, zenuwachtig of gespannen
- $\square_0$ $\square_1$ $\square_2$ 46. Zenuwachtige bewegingen of zenuwtrekken (schrijf op): ………………………………
- $\square_0$ $\square_1$ $\square_2$ 47. Voert regels te strikt uit, te aangepast
- $\square_0$ $\square_1$ $\square_2$ 48. Andere leerlingen mogen hem/haar niet
- $\square_0$ $\square_1$ $\square_2$ 49. Heeft moeilijkheden met leren
- $\square_0$ $\square_1$ $\square_2$ 50. Is te angstig of te bang
- $\square_0$ $\square_1$ $\square_2$ 51. Voelt zich duizelig of licht in het hoofd
- $\square_0$ $\square_1$ $\square_2$ 52. Voelt zich erg schuldig
- $\square_0$ $\square_1$ $\square_2$ 53. Praat voor zijn/haar beurt
- $\square_0$ $\square_1$ $\square_2$ 54. Voelt zich erg moe zonder reden
- $\square_0$ $\square_1$ $\square_2$ 55. Te dik
- 56. **Lichamelijke problemen zonder bekende medische oorzaak:**
  - $\square_0$ $\square_1$ $\square_2$ a. Pijnen (geen buikpijn of hoofdpijn)
  - $\square_0$ $\square_1$ $\square_2$ b. Hoofdpijn
  - $\square_0$ $\square_1$ $\square_2$ c. Misselijk
  - $\square_0$ $\square_1$ $\square_2$ d. Oogproblemen (waarvoor een bril of lenzen niet helpen) (schrijf op): ………………………………
  - $\square_0$ $\square_1$ $\square_2$ e. Huiduitslag of andere huidproblemen
  - $\square_0$ $\square_1$ $\square_2$ f. Buikpijn
  - $\square_0$ $\square_1$ $\square_2$ g. Overgeven
  - $\square_0$ $\square_1$ $\square_2$ h. Andere problemen (schrijf op): ………………………………

Graag met blokletters invullen. Graag alle vragen beantwoorden.

0 = Helemaal Niet (voor zover u weet) | 1 = Een Beetje of Soms | 2 = Duidelijk of Vaak

- ☐0 ☐1 ☐2 57. Valt mensen lichamelijk aan
- ☐0 ☐1 ☐2 58. Pulkt aan neus, huid of aan iets anders van het lichaam (schrijf op): ……
- ☐0 ☐1 ☐2 59. Slaapt tijdens de les
- ☐0 ☐1 ☐2 60. Onverschillig of ongemotiveerd
- ☐0 ☐1 ☐2 61. Schoolwerk is slecht
- ☐0 ☐1 ☐2 62. Onhandig of stuntelig
- ☐0 ☐1 ☐2 63. Gaat liever om met oudere jongens of meisjes
- ☐0 ☐1 ☐2 64. Gaat liever om met jongere jongens of meisjes
- ☐0 ☐1 ☐2 65. Weigert om te praten
- ☐0 ☐1 ☐2 66. Herhaalt bepaalde handelingen steeds maar weer, dwanghandelingen (schrijf op): ……
- ☐0 ☐1 ☐2 67. Veroorzaakt onrust in de klas
- ☐0 ☐1 ☐2 68. Schreeuwt veel
- ☐0 ☐1 ☐2 69. Gesloten, houdt dingen voor zichzelf
- ☐0 ☐1 ☐2 70. Ziet dingen die er niet zijn (schrijf op): ……
- ☐0 ☐1 ☐2 71. Schaamt zich gauw of voelt zich niet op zijn/haar gemak
- ☐0 ☐1 ☐2 72. Slordig werk
- ☐0 ☐1 ☐2 73. Gedraagt zich onverantwoordelijk (schrijf op): ……
- ☐0 ☐1 ☐2 74. Slooft zich uit of doet gek om op te vallen
- ☐0 ☐1 ☐2 75. Te verlegen of timide
- ☐0 ☐1 ☐2 76. Explosief en onvoorspelbaar gedrag
- ☐0 ☐1 ☐2 77. Wensen moeten onmiddellijk ingewilligd worden, snel gefrustreerd
- ☐0 ☐1 ☐2 78. Let niet goed op of is snel afgeleid
- ☐0 ☐1 ☐2 79. Spraakprobleem (schrijf op): ……
- ☐0 ☐1 ☐2 80. Kijkt met een lege blik
- ☐0 ☐1 ☐2 81. Kan niet tegen kritiek
- ☐0 ☐1 ☐2 82. Steelt
- ☐0 ☐1 ☐2 83. Spaart te veel dingen op die hij/zij niet nodig heeft (schrijf op): ……
- ☐0 ☐1 ☐2 84. Vreemd gedrag (schrijf op): ……
- ☐0 ☐1 ☐2 85. Vreemde gedachten (schrijf op): ……
- ☐0 ☐1 ☐2 86. Koppig, stuurs of prikkelbaar
- ☐0 ☐1 ☐2 87. Stemming of gevoelens veranderen plotseling
- ☐0 ☐1 ☐2 88. Mokt veel
- ☐0 ☐1 ☐2 89. Achterdochtig
- ☐0 ☐1 ☐2 90. Vloekt of gebruikt vieze woorden
- ☐0 ☐1 ☐2 91. Praat erover dat hij/zij zichzelf zou willen doden
- ☐0 ☐1 ☐2 92. Presteert beneden eigen niveau
- ☐0 ☐1 ☐2 93. Praat te veel
- ☐0 ☐1 ☐2 94. Pest veel
- ☐0 ☐1 ☐2 95. Driftbuien of snel driftig
- ☐0 ☐1 ☐2 96. Denkt te veel aan seks
- ☐0 ☐1 ☐2 97. Bedreigt mensen
- ☐0 ☐1 ☐2 98. Komt te laat op school of in de les
- ☐0 ☐1 ☐2 99. Rookt tabak
- ☐0 ☐1 ☐2 100. Voert opgedragen taken niet uit
- ☐0 ☐1 ☐2 101. Spijbelen of schoolverzuim zonder opgaaf van reden
- ☐0 ☐1 ☐2 102. Weinig actief, beweegt zich langzaam of te weinig energie
- ☐0 ☐1 ☐2 103. Ongelukkig, verdrietig of depressief
- ☐0 ☐1 ☐2 104. Meer dan gewoon luidruchtig
- ☐0 ☐1 ☐2 105. Gebruikt alcohol of drugs (schrijf op): ……
- ☐0 ☐1 ☐2 106. Wil te graag het de ander naar de zin maken
- ☐0 ☐1 ☐2 107. Heeft een hekel aan school
- ☐0 ☐1 ☐2 108. Is bang om fouten te maken
- ☐0 ☐1 ☐2 109. Zeuren
- ☐0 ☐1 ☐2 110. Onverzorgd uiterlijk
- ☐0 ☐1 ☐2 111. Teruggetrokken, gaat niet met anderen om
- ☐0 ☐1 ☐2 112. Maakt zich zorgen
- ☐0 ☐1 ☐2 113. Gedraagt zich als iemand van het andere geslacht
- ☐0 ☐1 ☐2 114. Eet of drinkt dingen die eigenlijk niet eet- of drinkbaar zijn, geen snoep (schrijf op): ……
- ☐0 ☐1 ☐2 115. Is teveel bezig met netjes of schoon zijn
- 116. Schrijf hier ieder ander probleem op dat deze leerling heeft en dat hierboven nog niet genoemd is:
  - ☐0 ☐1 ☐2 ……
  - ☐0 ☐1 ☐2 ……
  - ☐0 ☐1 ☐2 ……
  - ☐0 ☐1 ☐2 ……

APPENDIX C CONNERS' TEACHER RATINGS SCALE - REVISED

Conners' Leerkracht Beoordelingsschaal - Herziene Versie (S)

door C. Keith Conners, Ph.D.

Instructies: Hieronder staat een aantal vaak voorkomende problemen die kinderen op school kunnen hebben. Geef steeds aan in hoeverre het kind zich de afgelopen maand zo heeft gedragen. Stel uzelf bij elk probleem de vraag " In hoeverre is dit de afgelopen maand een probleem geweest?" en maak voor elk probleem het vakje bij het beste antwoord zwart. Maak het vakje met de 0 zwart als het antwoord is: niet, helemaal niet, zelden of bijna nooit. Maak het vakje met de 3 zwart als het antwoord is: helemaal waar, komt heel vaak voor of komt heel regelmatig voor. Maak het vakje met de 1 of 2 zwart als het antwoord er tussenin zit. Wij verzoeken u om <u>alle</u> vragen te beantwoorden.

0 = helemaal niet waar (nooit, zelden) **1 = klein beetje waar** (af en toe) **2 = behoorlijk waar** (vaak, regelmatig) **3 = helemaal waar** (heel vaak, heel regelmatig)

1.	Onoplettend, snel afgeleid	☐0	☐1	☐2	☐3
2.	Opstandig	☐0	☐1	☐2	☐3
3.	Onrustig in de zin van friemelig	☐0	☐1	☐2	☐3
4.	Vergeet dingen die hij/zij al geleerd heeft	☐0	☐1	☐2	☐3
5.	Stoort andere kinderen	☐0	☐1	☐2	☐3
6.	Wanneer een volwassene hem/haar iets vraagt weigert hij/zij dit of gaat er actief tegenin	☐0	☐1	☐2	☐3
7.	Is altijd "in de weer" of gedraagt zich zo dat het lijkt alsof hij/zij door een motor wordt aangedreven	☐0	☐1	☐2	☐3
8.	Slecht in spelling	☐0	☐1	☐2	☐3
9.	Kan niet rustig blijven	☐0	☐1	☐2	☐3
10.	Hatelijk of wraakzuchtig	☐0	☐1	☐2	☐3
11.	Staat op in de klas of in andere situaties waarin verwacht wordt dat je blijft zitten	☐0	☐1	☐2	☐3
12.	Friemelt met handen en voeten of wiebelt op zijn/haar stoel	☐0	☐1	☐2	☐3
13.	Loopt achter met lezen	☐0	☐1	☐2	☐3
14.	Houdt kort de aandacht bij dingen	☐0	☐1	☐2	☐3
15.	Spreekt volwassenen tegen	☐0	☐1	☐2	☐3
16.	Heeft alleen aandacht voor dingen die hem/haar interesseren	☐0	☐1	☐2	☐3
17.	Heeft moeite om op zijn/haar beurt te wachten	☐0	☐1	☐2	☐3
18.	Heeft weinig interesse in zijn/haar schoolwerk	☐0	☐1	☐2	☐3
19.	Snel afgeleid of heeft een probleem om de aandacht vast te houden	☐0	☐1	☐2	☐3
20.	Driftbuien: explosief, onvoorspelbaar gedrag	☐0	☐1	☐2	☐3
21.	Rent rond of klimt overal op in situaties waarin dit niet gepast is	☐0	☐1	☐2	☐3
22.	Is slecht in rekenen	☐0	☐1	☐2	☐3
23.	Onderbreekt of stoort anderen (bijv. mengt zich in gesprekken of spelletjes van anderen)	☐0	☐1	☐2	☐3
24.	Heeft moeite om tijdens spel of andere vrijetijdsactiviteiten rustig te zijn	☐0	☐1	☐2	☐3
25.	Maakt dingen niet af waar hij/zij mee begonnen is	☐0	☐1	☐2	☐3
26.	Volgt instructies niet op en maakt schoolwerk niet af (niet vanwege opstandig gedrag of het niet begrijpen van de instructies)	☐0	☐1	☐2	☐3
27.	Snel opgewonden, impulsief	☐0	☐1	☐2	☐3
28.	Onrustig, altijd in de weer	☐0	☐1	☐2	☐3

Vertaling 2001: Kinder- en Jeugdpsychiatrie, Erasmus Medisch Centrum Rotterdam

HARTELIJK DANK VOOR UW MEDEWERKING!

APPENDIX D EDUCATIONAL ACHIEVEMENT TEST FORM

UITSLAG CITO EINDTOETS BASISONDERWIJS

OUDSTE ***(eerstgeboren van de tweeling):***

Voornaam: ____________________ Geboortedatum ______-______-______

Heeft dit kind deelgenomen aan de Cito Eindtoets?
☐1 ja → vul onderdeel A in
☐2 nee → vul onderdeel B in

A Datum afname Cito-toets: ______-______-______

Resultaten (*vul in*):

	Taal	*Rekenen-Wiskunde*	*Studie-vaardigheden*	*Wereld-oriëntatie*	*Totaal*	*Standaard score*
Aantal opgaven:						
Aantal goed:						
Percentielscore:						

B De oudste van de tweeling heeft niet deelgenomen aan de Cito Eindtoets Basisonderwijs, want…

☐1 dit kind zit nog niet in groep 8
☐2 dit kind zit op het speciaal (basis)onderwijs
☐3 dit kind heeft een andere toets of test gedaan in groep 8, namelijk: ____________________
☐4 de school heeft de Cito Eindtoets Basisonderwijs wel afgenomen, maar dit kind heeft niet deelgenomen
☐5 de school neemt helemaal niet deel aan toetsen of testen
☐6 andere reden, namelijk: ____________________

JONGSTE ***(laatstgeborene van de tweeling):***

Voornaam: ____________________ Geboortedatum ______-______-______

Heeft dit kind deelgenomen aan de Cito Eindtoets?
☐1 Ja → vul onderdeel A in
☐2 nee → vul onderdeel B in

A Datum afname Cito-toets: ______-______-______

Resultaten (*vul in*):

	Taal	*Rekenen-Wiskunde*	*Studie-vaardigheden*	*Wereld-oriëntatie*	*Totaal*	*Standaard score*
Aantal opgaven:						
Aantal goed:						
Percentielscore:						

B De jongste van de tweeling heeft niet deelgenomen aan de Cito Eindtoets Basisonderwijs, want…

☐1 dit kind zit nog niet in groep 8
☐2 dit kind zit op het speciaal (basis)onderwijs
☐3 dit kind heeft een andere toets of test gedaan in groep 8, namelijk: ____________________
☐4 de school heeft de Cito Eindtoets Basisonderwijs wel afgenomen, maar dit kind heeft niet deelgenomen
☐5 de school neemt helemaal niet deel aan toetsen of testen
☐6 andere reden, namelijk: ____________________

APPENDIX E PUPIL MONITORING SYSTEM

NEDERLANDS - AVI

Toetsnaam	*Datum*	*Ind.*	*Onderdeel*	*Score*	*Schaal*	*DL*	*DLE*	*LR%*	*LR*	*Nw.*	*Niv.*
AVI-CITO	10-02-2011	Nee	AVI-Beheersing (0=Start t..	2	-	6	10	167	4	4,6	E3
AVI-CITO	09-06-2011	Nee	AVI-Beheersing (0=Start t..	3	-	10	15	150	5	4,4	M4
AVI-CITO	08-02-2012	Nee	AVI-Beheersing (0=Start t..	4	-	16	20	125	4	4,1	E4
AVI-CITO	18-06-2012	Nee	AVI-Beheersing (0=Start t..	4	-	20	20	100	0	3,0	E4
AVI-CITO	08-02-2013	Nee	AVI-Beheersing (0=Start t..	6	-	26	30	115	4	3,9	E5

NEDERLANDS - CITO Begrijpend lezen

Toetsnaam	*Datum*	*Ind.*	*Onderdeel*	*Score*	*Schaal*	*DL*	*DLE*	*LR%*	*LR*	*Nw.*	*Niv.*	
LOVS E3	14-06-2011	Nee	Start + Vervolg 2	36	3	10	12	120	2	3,5	B	II
LOVS M4	02-02-2012	Nee	Start + Vervolg 2	37	17	16	22	138	6	3,7	B	II
LOVS E4	25-06-2012	Nee	Start + Vervolg 2	32	15	20	21	105	1	3,0	B	III
LOVS M5	18-01-2013	Nee	Start + Vervolg 2	34	30	25	32	128	7	3,8	B	II

NEDERLANDS - CITO Drie-Minuten-Toets

Toetsnaam	*Datum*	*Ind.*	*Onderdeel*	*Score*	*Schaal*	*DL*	*DLE*	*LR%*	*LR*	*Nw.*	*Niv.*	
LOVS M3	09-02-2011	Nee	Kaart 1*	22	19	6	5	83	-1	2,1	C	IV
LOVS E3	04-07-2011	Nee	Kaart 1+2+3	130	43	10	12	120	2	3,7	B	II
LOVS M4	09-02-2012	Nee	Kaart 1*	71	62	16	19	119	3	3,2	B	III
			Kaart 2*	55	54	16	15	94	-1	2,9	C	III
			Kaart 3*	39	46	16	14	88	-2	3,1	B	III
			Kaart 1+2+3	165	55	16	15	94	-1	3,1	B	III
LOVS E4	04-07-2012	Nee	Kaart 1*	71	62	20	19	95	-1	2,6	C	III
			Kaart 2*	63	62	20	19	95	-1	2,8	C	III
			Kaart 3*	45	53	20	16	80	-4	2,8	C	III
			Kaart 1+2+3	179	59	20	17	85	-3	2,7	C	III
LOVS M5	23-01-2013	Nee	Kaart 1*	74	64	25	20	80	-5	2,3	C	IV
			Kaart 2*	75	74	25	26	104	1	3,0	B	III
			Kaart 3	63	73	25	26	104	1	3,0	B	III
			Kaart 1+2+3	212	70	25	23	92	-2	2,8	C	III

NEDERLANDS - CITO Spelling

Toetsnaam	*Datum*	*Ind.*	*Onderdeel*	*Score*	*Schaal*	*DL*	*DLE*	*LR%*	*LR*	*Nw.*	*Niv.*	
LOVS M3	07-02-2011	Nee	Start + Vervolg 2	48	115	6	12	200	6	4,3	A	I
LOVS E3	21-06-2011	Nee	Start + Vervolg 2	47	120	10	19	190	9	4,4	A	I
LOVS M4	02-02-2012	Nee	Start + Vervolg 2	40	121	16	20	125	4	3,6	B	II
LOVS E4	19-06-2012	Nee	Start + Vervolg 2	41	127	20	26	130	6	4,2	A	I
LOVS M5	23-01-2013	Nee	Start + Vervolg 2	35	127	25	26	104	1	3,3	B	III

NEDERLANDS - CITO Woordenschattoets

Toetsnaam	*Datum*	*Ind.*	*Onderdeel*	*Score*	*Schaal*	*DL*	*DLE*	*LR%*	*LR*	*Nw.*	*Niv.*	
LOVS M3	07-02-2011	Nee	Totaal	37	21	6	<6	<100	<0	1,0	D	V
LOVS E3	17-06-2011	Nee	Totaal	42	46	10	9	90	-1	2,7	C	III
LOVS M4	02-02-2012	Nee	Totaal	30	35	16	<6	<38	<-10	1,1	D	V
LOVS E4	19-06-2012	Nee	Totaal	40	55	20	17	85	-3	2,6	C	IV
LOVS M5	21-01-2013	Nee	Totaal	45	58	25	19	76	-6	2,4	C	IV

NEDERLANDS - CPS Beginnende Geletterdheid													
Toetsnaam	*Datum*	*Ind.*	*Onderdeel*	*Score*	*Schaal*	*DL*	*DLE*	*LR%*	*LR*	*Nw.*	*Niv.*	
Benoemsnelheid cijfers en	31-05-2010	Nee	Cijfers (sec)	46	-	-	-	-	-	2,0	C	IV
Letterkennistoets 2 (met gr		Nee	Letterkennis	13	-	-	-	-	-	2,3	C	
Rijmtoets (apr groep 1)		Nee	Rijmen	24	-	-	-	-	-	4,7	A	I
Synthesetoets 2 (met groep		Nee	Synthese	24	-	-	-	-	-	5,0	A	I

REKENEN/WISKUNDE - CITO Rekenen-Wiskunde												
Toetsnaam	*Datum*	*Ind.*	*Onderdeel*	*Score*	*Schaal*	*DL*	*DLE*	*LR%*	*LR*	*Nw.*	*Niv.*	
LOVS M3	07-02-2011	Nee	Totaal	40	33	6	9	150	3	3,7	B	II
LOVS E3	16-06-2011	Nee	Totaal	41	43	10	13	130	3	3,8	B	II
LOVS M4	03-02-2012	Nee	Totaal	45	60	16	22	138	6	4,1	A	I
LOVS E4	21-06-2012	Nee	Totaal	50	78	20	32	160	12	4,6	A+	I
LOVS M5	23-01-2013	Nee	Totaal	52	91	25	42	168	17	4,6	A+	I

APPENDIX F LETTER PARENTAL CONSENT

VU VRIJE UNIVERSITEIT AMSTERDAM

Nederlands Tweelingen Register

2500
Fam. «Achternaam»
«Straat» «Huisnummer» «Toevoeging»
«Postcode» «Plaatsnaam»

DATUM	ONS KENMERK	UW BRIEF VAN	UW KENMERK
december 2013	PC1 TRF«TRF»	NTR	«IDnummer»
E-MAIL	TELEFOON	FAX	BIJLAGE(N)
ntr.leerkrachten@vu.nl	020 598 8944	020 598 8832	-

Geachte ouder/verzorger,

U doet mee aan onderzoek van het Nederlands Tweelingen Register (NTR) naar de ontwikkeling van twee- en meerlingen. Dit onderzoek heeft, dankzij uw deelname en die van de duizenden andere meerlingfamilies, geleid tot belangrijke inzichten. Om een nog completer beeld te krijgen van de ontwikkeling van opgroeiende meerlingen betrekken we ook graag hun leerkrachten bij het onderzoek. Na uw toestemming, sturen wij de leerkrachten in de loop van het schooljaar (rond maart) een vragenlijst over gedrag en schoolprestaties.

Om de leerkrachten te kunnen benaderen, hebben wij uw toestemming en de gegevens van de school nodig. Als u naast de tweeling nog andere kinderen op de basisschool heeft dan willen wij hun leerkracht ook graag een vragenlijst sturen. Wilt u op onderstaande website aangeven of u toestemming geeft?

www.tweelingenregister.org/PC
Uw inlognaam is: «inlognaam»
Uw wachtwoord is: «wachtwoord»

Uw toestemming en de deelname van leerkrachten aan dit onderzoek is geheel vrijwillig. Alle gegevens worden vertrouwelijk behandeld. Dit betekent onder andere dat ouders, kinderen en leerkrachten geen inzage krijgen in elkaars antwoorden. Als u niet wilt dat wij de leerkracht van uw kinderen benaderen, vragen wij u dit ook aan te geven zodat u hierover geen post meer ontvangt.

In de afgelopen jaren hebben veel leerkrachten meegedaan. Hierdoor is bijv. duidelijk geworden dat het voor de sociale ontwikkeling van tweelingen niet uitmaakt of de kinderen wel of niet bij elkaar in de klas zitten. Meer hierover vindt u op: www.tweelingenregister.org/leerkrachten.

Als u liever niet via internet toestemming geeft, of als u vragen heeft, dan kunt u contact met ons opnemen per e-mail: ntr.leerkrachten@vu.nl of tel: 020-598 8944 (tijdens kantooruren).

Wij danken u bij voorbaat heel hartelijk voor uw medewerking!

Met vriendelijke groet,

Prof. dr. Dorret Boomsma

NEDERLANDS TWEELINGEN REGISTER (NTR)
www.tweelingenregister.org

BEZOEKADRES
Van der Boechorststraat 1
Transitorium

POSTADRES
Van der Boechorststraat 1
1081 BT Amsterdam

APPENDIX G PARENTAL CONSENT

Toestemmingsverklaring

Ik geef **wel / geen*** toestemming aan het Nederlands Tweelingen Register om de leerkracht(en) van onderstaande kinderen te benaderen ten behoeve van onderzoek naar de gedragsontwikkeling van kinderen.

Achternaam tweeling: ……………………………………………………………......

Geboortedatum tweeling: ……………………………………………………………..

Voorletters, achternaam ouder/verzorger: ………………… , ...…

Geslacht: man / vrouw*

Postcode: ………………………………………………………………........................

Plaats/datum: ………………………………………………………… ....

Handtekening: ……………………………………………………………..............

<u>Gegevens Tweeling</u>

Gegevens **oudste** van de tweeling *(eerstgeborene)*
Voornaam (roepnaam): ………………………… Geslacht: jongen/meisje* Naam school: ……………………………………………………………............ Postadres school (bij voorkeur postbusadres): ………….………………...... Postcode en Plaatsnaam school: ……………………………………………... Voornaam leerkracht: …………………………….... Geslacht: man / vrouw* Achternaam leerkracht: ……………....…………………..….……………...

Gegevens **jongste** van de tweeling *(laatstgeborene)*
Voornaam (roepnaam): ………………………… Geslacht: jongen/meisje* Naam school: ……………………………….………………………................ Postadres school (bij voorkeur postbusadres): …………………………….. Postcode en Plaatsnaam school: ……………………………………………... Voornaam leerkracht: …………………………….... Geslacht: man / vrouw* Achternaam leerkracht: ………...…………………….…………………….

** graag doorhalen wat niet van toepassing is* **Zie Achterzijde Formulier**

<u>Gegevens broertje(s) en/of zusje(s) van de tweeling</u>

Gegevens **broertje of zusje** van de tweeling

Voornaam (roepnaam): ………………………… Geslacht: jongen/meisje*
Achternaam: …………………………………………………………...............
Geboortedatum: ………………………………………………………..........
Naam school: …………………………………………………………..............
Postadres school (bij voorkeur postbusadres): …….………………….......
Postcode en Plaatsnaam school: ……………………………………………....
Voornaam leerkracht: …………………………....… Geslacht: man / vrouw*
Achternaam leerkracht: ……………...…………………...….……………....

Gegevens **broertje of zusje** van de tweeling

Voornaam (roepnaam): ………………………… Geslacht: jongen/meisje*
Achternaam: …………………………………………………………...............
Geboortedatum: ………………………………………………………..........
Naam school: …………………………………………………………..............
Postadres school (bij voorkeur postbusadres): …….………………….......
Postcode en Plaatsnaam school: ……………………………………………....
Voornaam leerkracht: …………………………....… Geslacht: man / vrouw*
Achternaam leerkracht: ……………...…………………...….……………....

Gegevens **broertje of zusje** van de tweeling

Voornaam (roepnaam): ………………………… Geslacht: jongen/meisje*
Achternaam: …………………………………………………………...............
Geboortedatum: ………………………………………………………..........
Naam school: …………………………………………………………..............
Postadres school (bij voorkeur postbusadres): …….………………….......
Postcode en Plaatsnaam school: ……………………………………………....
Voornaam leerkracht: …………………………....… Geslacht: man / vrouw*
Achternaam leerkracht: ……………...…………………...….……………....

APPENDIX H LETTER TEACHER SURVEY

VRIJE UNIVERSITEIT AMSTERDAM

Nederlands Tweelingen Register

2500
Fam. «Achternaam»
«Straat» «Huisnummer» «Toevoeging»
«Postcode» «Plaatsnaam»

DATUM	ONS KENMERK	UW BRIEF VAN	UW KENMERK
december 2013	PC1 TRF«TRF»	NTR	«IDnummer»
E-MAIL	TELEFOON	FAX	BIJLAGE(N)
ntr.leerkrachten@vu.nl	020 598 8944	020 598 8832	

Geachte «aanspreektitel» «achternaam»,

Het Nederlands Tweelingen Register (NTR) doet onderzoek naar de ontwikkeling van twee- en meerlingen en hun broertjes en zusjes. Hun ouders/verzorgers vullen vaak al sinds de geboorte van de kinderen vragenlijsten in over hun ontwikkeling. Een belangrijk aspect van het onderzoek betreft het gedrag van kinderen op school. De afgelopen jaren hebben veel leerkrachten meegedaan aan dit NTR onderzoek. Een aantal resultaten hiervan zijn voor u samengevat op onze website (www.tweelingenregister.org/leerkrachten). Daar vindt u ook meer informatie over het onderzoek.

Bij u in de klas zitten één of meerdere kinderen uit een meerlinggezin. Hun ouders hebben ons toestemming gegeven om u te benaderen. Wij willen u vragen of u een vragenlijst wilt invullen over **«voornaamll» «achternaamll»**. Dit kunt u doen op:

www.tweelingenregister.org/TRF
Uw inlognaam is: «inlognaam»
Uw wachtwoord is: «wachtwoord»

Alle gegevens worden vertrouwelijk behandeld. Dit betekent onder meer dat ouders geen inzage krijgen in de vragenlijst ingevuld door de leerkracht en dat de leerkracht geen inzage krijgt in de antwoorden van ouders.

Als uw school de leerlingvolgtoetsen van het Cito afneemt, willen wij u vragen om het leerlingrapport (**o.v.v. «IDNummer»**) als bijlage te mailen naar ntr.leerkrachten@vu.nl of als uitdraai op te sturen naar: NTR, Antwoordnummer 2941, 1000 SN Amsterdam (een postzegel is niet nodig).

Als de leerlingadministratie wordt bijgehouden in het Cito computerprogramma LOVS, verzoeken wij u om bij het aanmaken van het rapport te kiezen voor de weergave van de toetsresultaten in een tabel. Wij ontvangen graag de (beschikbare) resultaten van de toetsen **Taal voor Kleuters**, **Rekenen voor Kleuters**, **DMT**, **Rekenen-Wiskunde**, **Spelling** en **Begrijpend Lezen** van de afnamemomenten halverwege het schooljaar (**M1 t/m M8**).

We realiseren ons dat leerkrachten het erg druk hebben, maar stellen uw medewerking bijzonder op prijs! Mocht u vragen hebben of liever een papieren vragenlijst invullen, dan kunt u contact met ons opnemen via e-mail: ntr.leerkrachten@vu.nl of via telefoon: 020 - 598 8944.

Met vriendelijke groet en bij voorbaat heel hartelijk dank voor uw medewerking,

Mevr. Prof. dr. Dorret Boomsma

NEDERLANDS TWEELINGEN REGISTER (NTR)	**BEZOEKADRES**	**POSTADRES**
www.tweelingenregister.org	Van der Boechorststraat 1 Transitorium	Van der Boechorststraat 1 1081 BT Amsterdam

APPENDIX I INFORMATION ON WEBSITE

Een kind brengt veel tijd op school door. Daarom stuurt het Nederlands Tweelingen Register, na toestemming van de ouders, ook vragenlijsten naar de leerkrachten van meerlingen. Met informatie van ouders/verzorgers en leerkrachten ontstaat zo een breed beeld van de gedragsontwikkeling van een kind. Naast informatie over het gedrag van kinderen verzamelen wij ook gegevens over de cognitieve ontwikkeling en schoolprestaties van kinderen. Bovendien vragen we leerkrachten om de ontwikkeling van broertjes en zusjes van meerlingen te beoordelen. Dit is belangrijk omdat er nog steeds vragen zijn over een mogelijk vertraagde ontwikkeling van meerlingen. Via deze pagina willen wij u informeren over enkele onderzoeken en resultaten die zijn verkregen met de hulp van de leerkrachten. Het Nederlands Tweelingen Register hoopt in de toekomst nog meer gegevens te verzamelen van leerkrachten om steeds meer te weten komen over de ontwikkeling van kinderen.

Tweelingen niet in dezelfde klas? Onzin!

Een terugkerende vraag voor ouders van meerlingen is of het voor de ontwikkeling van de kinderen beter is om op de basisschool in dezelfde klas of in aparte klassen te zitten. Tinca Polderman en Marieke van Leeuwen hebben onderzocht of gedragsproblemen en de scores op de Cito-toets van tweelingen samen in de klas verschillen van die van tweelingen die gescheiden les krijgen. Op 7- en 12-jarige leeftijd blijken tweelingen die gescheiden les krijgen volgens leerkrachten iets meer gedragsproblemen te vertonen dan tweelingen die samen in een klas zitten. Dit verschil komt echter meestal door reeds bestaande gedragsproblemen en niet door het scheiden van de kinderen. Voor de scores op de Cito-toets blijkt het niet uit te maken uit of tweelingen wel of niet samen in de klas hebben gezeten. Of tweelingen samen in de klas kunnen of niet kan het beste door ouders en school in overleg bepaald worden. Op grond van onderzoek is het een niet beter dan het ander.

Tweelingen en hun schoolprestaties

In een recent onderzoek van Eveline de Zeeuw onderzocht zij de invloed van een aantal risicofactoren van een tweelinggeboorte op schoolprestaties (rekenen, taal, lezen en gym). Daarnaast werd in dit onderzoek gekeken of tweelingen hetzelfde scoorden op school als hun broers en zussen. De resultaten lieten zien dat hoe lager het geboortegewicht hoe slechter de beoordelingen op de verschillende inhoudelijke schoolvakken. Hoe goed tweelingen waren in gym werd vooral beïnvloed door of er complicaties waren opgetreden na de geboorte en hoe lang ze in een couveuse hadden gelegen. Tweelingen bleken lagere cijfers te krijgen op de schoolvakken in vergelijking met hun oudere broers en zussen maar niet in vergelijking met hun jongere broers en zussen. Dit betekent dat de geboortevolgorde in een gezin en niet de risicofactoren van de tweelinggeboorte (een deel van) het verschil in schoolprestaties lijkt te bepalen.

Genetische invloeden van belang bij gedragsproblemen op school

Waarom vertoont het ene kind meer gedragsproblemen dan het andere kind? Onderzoek bij tweelingen geeft inzicht in de mate waarin gedrag wordt bepaald door erfelijke aanleg of door de omgeving. Eske Derks heeft onderzocht waar verschillen tussen kinderen in aandachtsproblemen op school vandaan komen. Uit de vele ingevulde vragenlijsten bleek dat verschillen in aandachtsproblemen voor een groot deel bepaald worden door genetische factoren. Soortgelijk onderzoek van Tinca Polderman heeft aangetoond dat ook verschillen in angstig, depressief of teruggetrokken gedrag en verschillen in opstandig, agressief of normafwijkend gedrag voor een aanzienlijk deel verklaard kunnen worden door genetische invloeden.

De feminisering van het basisonderwijs

In de media en politiek wordt beweerd dat het toenemende aantal vrouwelijke leerkrachten op de basisschool een slechte invloed heeft op de schoolprestaties en het gedrag van leerlingen, met name bij jongens. Eveline de Zeeuw heeft in twee unieke groepen kinderen gekeken of er bewijs te vinden was voor deze bewering. Dit heeft ze gedaan door te kijken naar een groep eeneiige tweelingen waarvan het ene kind bij een meester en het andere kind bij een juf in de klas zat en naar een groep jongen-meisje tweelingen die allebei of bij een juf of bij een meester in de klas zaten. Uit dit onderzoek is gebleken dat een leerkracht van hetzelfde geslacht geen invloed heeft op de schoolprestaties of de aanwezigheid van gedragsproblemen van jongens of meisjes.

Narcose geen oorzaak van leerproblemen

Uit een onderzoek van Meike Bartels blijkt dat het ondergaan van een narcose op jonge leeftijd geen oorzaak is van eventuele latere leerproblemen. Kinderen die op jonge leeftijd onder narcose zijn geweest blijken wel lagere scores op de Cito-toets te hebben dan kinderen die nog nooit onder narcose zijn geweest. Echter, eeneiige tweelingparen waarvan de het ene kind wel en het andere niet onder narcose is geweest verschillen niet in hun scores op de Cito-toets. Hieruit blijkt dat het onder narcose zijn geweest niet de oorzaak is van de lagere scores en dat er andere factoren van invloed zijn op de schoolprestaties.

Borstvoeding of flesvoeding?

Meike Bartels heeft ook onderzocht of het krijgen van borstvoeding invloed heeft op schoolprestaties. Kinderen die borstvoeding hebben gehad scoren ongeveer 1 tot 2 punten hoger op de Cito-toets dan flesgevoede kinderen. Ter vergelijking, kinderen van hoogopgeleide moeders scoren ongeveer 6 tot 7 punten hoger op deze toets dan kinderen van laagopgeleide moeders. Daarnaast geven hoogopgeleide moeders vaker en langer borstvoeding. Hoewel het effect van borstvoeding ten opzichte van het effect van opleidingsniveau klein is, hebben kinderen van hoogopgeleide moeders die de borst krijgen wel een dubbel voordeel.

LIST OF PUBLICATIONS

ARTICLES

de Zeeuw, E.L., Tak, E.C.P.M., Dusseldorp, E. & Hendriksen, I.J.M. (2010). Workplace exercise intervention to prevent depression: A pilot randomized controlled trial. *Mental Health and Physical Activity, 3,* p. 72-77.

de Zeeuw, E.L., van Beijsterveldt, C.E.M., de Geus, E.J.C. & Boomsma, D.I. (2010). Twin specific risk factors in primary school achievements. *Twin Research and Human Genetics, 15 (1)*, p. 107-115.

van Beijsterveldt, C.E.M., Groen-Blokhuis, M., Hottenga, J.J., Franić, S., Hudziak, J.J., Lamb, D., Huppertz, C., de Zeeuw, E.L., Nivard, M., Schutte N., Swagerman, S., Glasner, T., van Fulpen M., Brouwer C., Stroet, T., Nowotny, D., Ehli, E.A., Davies, G.E., Scheet, P., Orlebeke J.F., Kan, K-J., Smit, D., Dolan, C.V., Middeldorp, C.M., de Geus, E.J.C., Bartels, M. & Boomsma D.I. (2013). The Young Netherlands Twin Register (YNTR): Longitudinal twin and family studies in over 70,000 children. *Twin Research and Human Genetics, 16 (1),* p. 252-267.

de Zeeuw, E.L., van Beijsterveldt, C.E.M., Glasner, T.J., Bartels, M., Ehli, E.A., Davies, G.E., Hudziak, J.J., Consortium SSGA, Rietveld, C.A., Groen-Blokhuis, M.M., Hottenga, J.J., de Geus, E.J.C. & Boomsma, D.I. (2014). Polygenic scores associated with educational attainment in adults predict educational achievement and ADHD symptoms in children. *American Journal of Medical Genetics Part B. 9999,* p. 1-11.

de Zeeuw, E.L., van Beijsterveldt, C.E.M., Glasner, T.J., de Geus, E.J.C. & Boomsma, D.I. (2014). Do children perform and behave better at school when taught by same-gender teachers? *Learning and Individual Differences, 36,* p. 152-156.

de Zeeuw, E.L., van Beijsterveldt, C.E.M., Glasner, T.J. & Boomsma, D.I. Childhood ODD and ADHD behavior: Effects of gender, teacher, genes and their interactions. *Behavioral Genetics*. Accepted for publication.

de Zeeuw, E.L., van Beijsterveldt, C.E.M., Glasner, T.J., de Geus, E.J.C. & Boomsma, D.I. (2015). Arithmetic, reading and writing performance has a strong genetic component: A study in primary school children. In Preparation.

de Zeeuw, E.L., de Geus, E.J.C. & Boomsma, D.I. (2015). A review and meta-analysis of twin studies on educational achievement in primary school highlights the importance of genetic variation. *Trends in Neuoroscience and Education.* Submitted.

de Zeeuw, E.L., van Beijsterveldt, C.E.M., Bartels, M., Middeldorp, C.M., de Geus, E.J.C. & Boomsma, D.I. (2015). Longitudinal genetic causal modeling suggests that ADHD symptoms lead to lower educational achievement in children. In Preparation.

PUBLISHED ABSTRACTS

de Zeeuw, E.L., van Beijsterveldt, C.E.M. & Boomsma, D.I. (2012). Do boys with a male teacher perform and behave better at school? A discordant monozygotic twin design. *Behavioral Genetics, 42 (6)*, p. 929.

de Zeeuw, E.L., van Beijsterveldt, C.E.M., Glasner, T.J., de Geus, E.J.C. & Boomsma, D.I. (2013). Teacher-rated educational achievement in twins in primary school. *Behavioral Genetics, 43 (6),* p. 514.

BOOK CHAPTERS

de Zeeuw, E.L., van Beijsterveldt, C.E.M. & Boomsma, D.I. (2012). Waarom de appel vaak niet ver van de boom valt. De invloed van genen en omgeving op de ontwikkeling van een kind. In: Kinderen-in-ontwikkeling op de basisschool. Antwerpen-Apeldoorn: Garant.

BOOK REVIEWS

de Zeeuw, E.L. (2014). Book Review: G is for Genes: The impact of genetics on education and achievement. *Twin Research and Human Genetics, 17 (2)*, p. 140-141.

DANKWOORD

DANKWOORD

De afgelopen jaren lijken voorbij gevlogen. Opeens is het zo ver: ik heb mijn proefschrift af. Dit had ik niet voor elkaar gekregen zonder de hulp en steun van een aantal mensen die ik daarom graag wil bedanken.

Zonder de medewerking van de duizenden tweelingen en hun ouders en leerkrachten zou dit onderzoek niet mogelijk zijn geweest en daarom wil ik hen hartelijk bedanken voor het invullen van alle, af en toe behoorlijk lange, vragenlijsten.

Mijn promotoren en co-promotor wil ik graag bedanken voor hun goede begeleiding. Dorret, jouw onuitputtelijke passie voor de wetenschap is inspirerend. De snelheid waarmee jij uitgebreide feedback geeft op ingeleverde stukken is onvoorstelbaar. Eco, jouw vermogen om, ogenschijnlijk moeiteloos, teksten te schrijven, is geweldig. Toos, de nauwkeurigheid en grondigheid waarmee jij alle data checkt, is bewonderenswaardig. Ik heb veel van je geleerd over de dataverzameling, SPSS syntax en statistische analyses.

Graag wil ik ook de leescommissie, prof. dr. Pak Sham, prof. dr. Jelle Jolles, prof. dr. Peter de Jong, prof. dr. Dinand Webbink, prof. dr. Conor Dolan en dr. Marjolein Rietveld, bedanken voor de tijd en aandacht die ze hebben besteed aan het lezen van mijn proefschrift.

Het Nederlands Tweelingen Register kan niet zonder de medewerkers van het YNTR, de collega's van het secretariaat en natuurlijk Natascha. Als er iets geregeld moet worden dan zijn zij altijd bereid om te helpen. Tina, bedankt voor al je hulp bij de dataverzameling van de TRF. Samen hebben we duizenden vragenlijsten verstuurd. Speciale dank aan alle aio's, oud en nieuw, voor de gezelligheid tijdens congresbezoeken, Boulder, de lunch en de aio club. En last maar zeker niet least, 2C33-S. Charlotte, Nienke en Ineke (en Suzanne), ik had me geen betere kamergenootjes kunnen wensen! Onze kamer is in de loop van de jaren, door alle planten en spullen aan de muur, getransformeerd tot de kleurrijkste kamer van de afdeling, en ik heb er een paar vriendinnen bij gekregen. Niet alleen op werk, maar ook tijdens roadtrips, etentjes, kerstmarkt, klimmen, pilates en zingen.

Een proefschrift kan niet geschreven worden zonder af en toe pauze te nemen. Voor de nodige ontspanning met kopjes thee, volleybal, etentjes, spelletjes, biertjes, skaten en andere activiteiten, kan ik altijd rekenen op mijn vriendinnen, vrienden en teamgenootjes. Ik ga hier verder geen namen noemen, maar ik wil jullie allemaal heel erg bedanken! Linda, wat hebben wij samen een mooie, verre reizen gemaakt en een hoop dingen beleefd. Ook dit jaar staan er weer bijzondere momenten op de planning. Vere, inmiddels ken ik je al meer jaren wel dan niet. Nu ga ik net als jij promoveren aan de VU. We lijken elkaar overal te volgen en ik hoop dat dat in de toekomst ook zo blijft. Ik ben blij dat jullie straks als paranimfen naast me willen staan om me door de verdediging van mijn proefschrift heen te slepen.

En ten slotte, mijn lieve familie, mama en José, en oma en de rest van mijn kleine, gezellige familie. Jullie geloofden dat ik dit tot een goed einde kon brengen en zijn altijd geïnteresseerd in waar ik mee bezig ben. Dank voor alle steun, ook in de moeilijke tijden. José, je hebt zelfs nog een tijdje op dezelfde afdeling gewerkt om te helpen met het invoeren van data voor mijn project. Mama, ik vind het altijd leuk om je verhalen uit de praktijk te horen. Jij bent voor mij het voorbeeld van een goede en betrokken lerares. Papa, wat had ik graag gewild dat je hier bij kon zijn. Jij was degene die riep dat ik moest gaan promoveren. Dankjewel! Ik weet zeker dat je trots op me zou zijn.